Statistical Methods for Cooperative and Distributed Inference in Wireless Networks

Politecnico di Torino

Scuola Interpolitecnica di Dottorato
Doctoral Program in Electronics and Communications

Final Dissertation

Statistical Methods for Cooperative and Distributed Inference in Wireless Networks

Federico Penna

Advisor
Prof. Roberto Garello

Coordinator of the Research Doctorate Course
Prof. Ivo Montrosset

March 2012

Federico Penna: *Statistical Methods for Cooperative and Distributed Inference in Wireless Networks*
Politecnico di Torino, © March 2012

ACKNOWLEDGMENTS

This thesis owes a lot to many people I had the chance to work with, in Torino and all over the world.

My deepest gratitude goes to my advisor, Prof. Roberto Garello, for the motivation he gave me and for his invaluable professional and human support during all my Ph.D.

Then, I would like to thank Dr. Maurizio Spirito for hosting me at the Pervasive Technologies Research Lab of Istituto Superiore Mario Boella (ISMB), as well as my colleagues Dr. Mauricio Caceres, Dr. Hussein Khaleel, Claudio Pastrone, Riccardo Tomasi, Francesco Sottile, Dr. Mirko Franceschinis, for their fruitful collaboration and all the time spent together. Chapter 4 of this thesis contains the main results of my work with ISMB.

I am greatly indebted to Prof. Henk Wymeersch for all that I learned from him during my three-month visit at Chalmers University in 2010. The work presented in Chapter 6 was done during this period, together with Mauricio Caceres. Chapter 9 is also fruit of my collaboration with Prof. Henk Wymeersch and with Vladimir Savic (Universidad Politecnica de Madrid, Spain).

I am grateful to Prof. Danijela Cabric, who gave me the great opportunity to spend six months at UCLA in 2010-11. The work presented in Chapter 8 of this thesis was done during my visit at UCLA, in collaboration with Yifan Sun, Prof. Lara Dolecek, and Prof. Danijela Cabric.

Special thanks to all my colleagues both at Chalmers and at UCLA for the friendly and welcoming atmosphere that really made me feel at home. I also wish to thank Prof. Boaz Nadler (Weizmann Institute, Israel) for his contribution to the work presented in Chapter 3. Finally, I would like to thank the Scuola Interpolitecnica di Dottorato (SIPD) for generously funding my research visits at Chalmers and UCLA.

ABSTRACT

This thesis is concerned with application of statistical methods – namely, random matrix theory (RMT) and belief propagation (BP) – in distributed inference problems in wireless communication networks. The term "distributed inference" denotes, in general, detection/estimation involving multiple network nodes ("sensors") that collect physical measurements and communicate with each other. Such problems can be classified as homogeneous, where all nodes observe the same hidden variable, or heterogeneous, where a different hidden variable exists for each node.

The first part of the thesis focuses on a homogeneous inference problem, i.e., multi-sensor signal detection in cognitive radio networks. Techniques based on the eigenvalues of sample covariance matrices are employed. The performance of such methods is analyzed mathematically, by using RMT results.

The second part addresses several heterogeneous inference problems: 1. distributed localization (in a hybrid scenario with GPS and terrestrial range measurements); 2. signal detection in non-uniform radio environments; 3. cooperative signal detection in the presence of malicious users. For each of these problems, BP-based Bayesian inference methods are adopted. In cases 1 and 2, in particular, the BP algorithm is implemented in a decentralized fashion, by exploiting the correspondence between statistical graph and physical network structure. Finally, a variation of the BP algorithm is proposed, providing improved performance in the case of graphs containing cycles.

PUBLICATIONS

Some of the contents of this thesis have previously appeared or will appear in the following publications:

Book chapters

1. [97] **F. Penna**, C. Pastrone, H. Khaleel, M. A. Spirito, R. Garello, "Characterization of Wi-Fi interference for dynamic channel allocation in WPANs", *Reliable Communications for Short-Range Wireless Systems* (edited by I. Guvenc et al.), Cambridge University Press, 2011.

Journal papers

1. [98] **F. Penna**, R. Garello, "Decentralized Neyman-Pearson Test with Belief Propagation for Peer-to-peer Collaborative Spectrum Sensing", accepted for publication in *IEEE Transactions on Wireless Communications*, 2012 (to appear).

2. [159] H. Wymeersch, **F. Penna**, V. Savic, "Uniformly Reweighted Belief Propagation for Distributed Inference", accepted for publication in *IEEE Transactions on Wireless Communications*, 2012 (to appear).

3. [57] H. Khaleel, **F. Penna**, C. Pastrone, R. Tomasi, M. A. Spirito, "Frequency-Agile Wireless Sensor Networks: Design and Implementation", accepted for publication in *IEEE Sensors Journal*, 2012 (to appear).

4. [99] **F. Penna**, Y. Sun, L. Dolecek, D. Cabric, "Detecting and Counteracting Statistical Attacks in Cooperative Spectrum Sensing", *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1806–1822, Apr. 2012.

5. [16] M. Caceres, **F. Penna**, H. Wymeersch, R. Garello, "Hybrid Cooperative Positioning based on Distributed Belief Propagation", *IEEE Journal on Selected Areas of Communications*, vol. 29, no. 10, pp. 1948–1958, Dec. 2011.

6. [100] **F. Penna**, R. Garello, "Detection of Discontinuous Signals for Cognitive Radio Applications", *IET Communications*, vol. 5, no. 10, pp. 1453–1561, Aug. 2011.

7. [101] **F. Penna**, M. A. Caceres, H. Wymeersch, "Cramér-Rao Bound for Hybrid GNSS-Terrestrial Cooperative Positioning", *IEEE Communications Letters*, vol. 14, no.11, pp. 1005–1007, November 2010.

8. [102] **F. Penna**, R. Garello, M. A. Spirito, "Cooperative Spectrum Sensing based on the Limiting Eigenvalue Ratio Distribution in Wishart Matrices", *IEEE Communications Letters*, vol.13, no.7, pp.507–509, July 2009.

Conference papers

1. [103] **F. Penna**, D. Cabric, "Bounds and Tradeoffs for Cooperative DoA-only Localization of Primary Users", *IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011.

2. [104] **F. Penna**, Y. Sun, L. Dolecek, D. Cabric, "Joint Spectrum Sensing and Detection of Malicious Nodes via Belief Propagation", *IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011.

3. [160] H. Wymeersch, **F. Penna**, V. Savic, "Uniformly Reweighted Belief Propagation: A Factor Graph Approach", *IEEE International Symposium on Information Theory (ISIT)*, Saint Petersburg, Russia, July 2011.

4. [105] **F. Penna**, J. Wang, D. Cabric, "Cooperative localization of primary users by directional antennas or antenna arrays: Challenges and design issues", *IEEE International Symposium on Antennas and Propagation (APS-URSI)*, Spokane, WA, USA, July 2011.

5. [106] **F. Penna**, H. Wymeersch, V. Savic, "Uniformly Reweighted Belief Propagation for Distributed Bayesian Hypothesis Testing", *IEEE Statistical Signal Processing Workshop (SSP)*, Nice, France, June 2011.

6. [88] B. Nadler, **F. Penna**, R. Garello, "Performance of Eigenvalue-based Signal Detectors with Known and Unknown Noise Level", *IEEE ICC*, Kyoto, Japan, June 2011.

7. [124] V. Savic, H. Wymeersch, **F. Penna**, S. Zazo, "Optimized Edge Appearance Probabilities for Cooperative Localization based on Non-parametric Tree-Reweighted Belief Propagation", *IEEE ICASSP*, Prague, Czech Republic, May 2011.

8. [107] **F. Penna**, R. Garello, M. A. Spirito, "Likelihood-Ratio Propagation and Consensus in Wireless Networks with Markov Random Field Models" *IEEE GLOBECOM - Workshop on Mobile Computing and Emerging Communication Networks*, Miami, FL, USA, Dec. 6-10, 2010

9. [14] M. Caceres, **F. Penna**, H. Wymeersch, R. Garello, "Hybrid GNSS-Terrestrial Positioning via Distributed Belief Propagation", *IEEE GLOBECOM 2010*, Miami, FL, USA, Dec. 6-10, 2010

10. [108] **F. Penna**, R. Tomasi, H. Khaleel, C. Pastrone, M. Spirito, "On Spectrum Sensing Duration in Cognitive Wireless Sensor Networks", *3rd*

International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART) (invited), Rome, Italy, Nov. 8-10, 2010

11. [135] R. Tomasi, H. Khaleel, **F. Penna**, C. Pastrone, M. Spirito, R. Garello, "Frequency Agility in IPv6-Based Wireless Personal Area Networks (6LoWPAN)" (invited paper), *Wired/Wireless Internet Communications (WWIC)*, Lulea, Sweden, June 1-3, 2010, in *Springer-Verlag Lecture Notes in Computer Science*, vol. 6074/2010, pp. 146-157

12. [109] **F. Penna**, R. Garello, M. A. Spirito, "Distributed Inference of Channel Occupation Probabilities in Cognitive Networks via Message Passing", *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Singapore, April 6-9, 2010

13. [56] H. Khaleel, **F. Penna**, C. Pastrone, R. Tomasi, M. Spirito, "Distributed Spectrum Sensing and Channel Selection in Opportunistic Wireless Personal Area Networks", *ACM MobiOpp 2010 - Demo Session*, Pisa, Italy, Feb. 22-24, 2010 - ***Best Demo Award***

14. [110] **F. Penna**, R. Garello, M. A. Spirito, "Probability of Missed Detection in Eigenvalue Ratio Spectrum Sensing", *5th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Marrakech, Morocco, Oct. 12-14, 2009

15. [55] H. Khaleel, C. Pastrone, **F. Penna**, M. A. Spirito, R. Garello, "Impact of Wi-Fi Traffic on the IEEE 802.15.4 Channels Occupation in Indoor Environments", *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Turin, Italy, Sept. 14-18, 2009

16. [111] **F. Penna**, R. Garello, D. Figlioli, M. A. Spirito, "Exact Non-asymptotic Threshold for Eigenvalue-based Spectrum Sensing", *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, Hannover, Germany, Jun. 22-24, 2009

17. [112] **F. Penna**, C. Pastrone, M. A. Spirito, R. Garello, "Energy Detection Spectrum Sensing with Discontinuous Primary User Signal", *IEEE International Conference on Communications (ICC)*, Dresden, Germany, Jun. 14-18, 2009

18. [113] **F. Penna**, C. Pastrone, M. A. Spirito, R. Garello, "Measurement-based Analysis of Spectrum Sensing in Adaptive WSNs under Wi-Fi and Bluetooth Interference", *IEEE 69th Vehicular Technology Conference (VTC)*, Barcelona, Spain, Apr. 26-29, 2009

19. [114] **F. Penna**, C. Pastrone, M. A. Spirito, R. Garello, "An Experimental Study on Spectrum Sensing for Cognitive WSNs under Wi-Fi Interference", *21st Wireless World Research Forum (WWRF) meeting*, Stockholm, Sweden, Oct. 13-15, 2008

CONTENTS

ACRONYMS

AWGN additive white gaussian noise

BER bit error rate

BP belief propagation

CDF cumulative distribution function

CED cooperative energy detection

CH Cluster Head

CLT central limit theorem

CR cognitive radio

CRB Cramér-Rao bound

CSCG circularly symmetric complex Gaussian

DSA dynamic spectrum allocation

DVB digital video broadcasting

EAP edge appearance probability

ECEF Earth-centered, Earth-fixed

ED energy detection

EDGE Enhanced Data Rates for GSM Evolution

ENU easting-northing-up

ERD eigenvalue-ratio detector

FA frequency agility

FA-WSN frequency-agile wireless sensor network

FAM Frequency Agility Manager

FC fusion center

FCC Federal Communications Commission

FG factor graph

FIM Fisher information matrix

FLOP floating point operations per second

GLRT generalized likelihood ratio test

GNSS global navigation satellite systems

GPS Global Positioning System

GSM Global System for Mobile Communications

H-SPAWN hybrid sum-Product algorithm over a wireless network

HC-EKF hybrid-cooperative extended Kalman filter

HC-WLS hybrid-cooperative weighted least squares

HRLRT Hybrid Roy's largest root test

HSPA High-Speed Packet Access

i.i.d. independent identically distributed

IoT Internet of things

IP Internet protocol

IPv6 Internet protocol version 6

ISM industrial, scientific, and medical

KLD Kullback-Leibler divergence

LDPC low-density parity-check

LLR log-likelihood ratio

LoS line-of-sight

LBP loopy belief propagation

LRT likelihood ratio test

MAC media access control

MAP maximum a posteriori

ML maximum likelihood

MMSE minimum mean squared error

MRF Markov random field

NBP nonparametric belief propagation

NP Neyman-Pearson

OC operating channel

P2P peer-to-peer

PDF probability distribution function

PHY physical [layer]

PSK Phase Shift Keying

PU primary user

QAM Quadrature Amplitude Modulation

RF radio-frequency

RCSS reputation-based cooperative spectrum sensing

RLRT Roy's largest root test

RMT random matrix theory

RMSE root mean squared error

ROC receiver operating characteristic

RSSI received signal strength indicator

SNR signal-to-noise ratio

SPA sum-product algorithm

SSN Spectrum Sensing Node

SU secondary user

TCP Transmission Control Protocol

TRW-BP tree-reweighted belief propagation

UAV unmanned aerial vehicle

UDP User Datagram Protocol

UMTS Universal Mobile Telecommunications System

URW-BP uniformly reweighted belief propagation

URW-NBP uniformly reweighted nonparametric belief propagation

UWB Ultra Wide-Band

WiMAX Worldwide Interoperability for Microwave Access

WPAN wireless personal area network

WRAN wireless regional area network

WSN wireless sensor network

Part I

BACKGROUND

1

DETECTION AND ESTIMATION IN WIRELESS NETWORKS

1.1 INTRODUCTION AND DEFINITIONS

Over the past few decades, voice and data communications have been revolutionized by the diffusion of a variety of new wireless technologies: mobile communication systems (GSM, EDGE, HSPA, UMTS), wireless Internet (Wi-Fi – IEEE 802.11, WiMAX – IEEE 802.16), low-power networks (wireless sensor networks (WSNs) and wireless personal area networks (WPANs) – IEEE 802.15.4, Zigbee), satellite navigation systems (Global Positioning System (GPS), Galileo), and the upcoming cognitive radio (CR) paradigm (first adopted in wireless regional area networks (WRANs) – IEEE 802.22).

From a signal processing perspective, several aspects of the operation of the above systems can be formulated as **distributed inference problems**. *Inference* denotes in general *detection* or *estimation* of some hidden variables, based on a set of observable data. Inference problems, when applied to wireless networks, are *distributed* in that they are performed cooperatively by a number of intercommunicating wireless devices (often referred to as *sensors*).

The following terminology can be used to classify distributed inference problems.

- According to the problem configuration: if all sensors observe the same hidden variable, inference problems are defined *homogeneous* or *global*; if different sensors may observe different, localized events, inference problems are referred to as *heterogeneous* or *local*.
- According to the architecture: distributed systems may be *centralized* (if there is a fusion center which processes the data sent from the other sensors), or *decentralized* (if all sensors behave as peers and work without a central control unit)[1].

In this thesis statistical methods are used to address both the aforementioned types of problems. In particular, **random matrix theory** will be applied to the design and performance evaluation of *homogeneous detection problems*, where a covariance matrix can be constructed from signal samples received by multiple sensors (assuming a *centralized architecture*). Then,

1 Note that this terminology is not universal: in the literature sometimes the word "distributed" is used as a synonym of "decentralized". To avoid any ambiguity, in this thesis the term "distributed" is used with a general meaning of "multi-sensor", and then a distinction is made between centralized and decentralized architectures.

Problem	*Goal*	*Observations*	*Type of hidden variable*
Signal detection in noise	Test hypothesis $\mathcal{H} = 1$ (signal present) vs. $\mathcal{H} = 0$ (no signal)	Received signal samples	Binary
(Cooperative) localization	Estimate the position of a wireless device in a network	Peer-to-peer or satellite (GPS) distance measurements	Continuous, 2-D or 3-D
Surveillance systems / sensor networks	Infer the presence of an object or event in a certain area	Physical measurements (temperature, light etc.)	Discrete

Table 1.1: Examples of distributed inference problems.

belief propagation will be used to address *heterogeneous detection and estimation problems*, where statistical dependencies among variables of the problem are represented by means of probabilistic graphical models. In some cases, the graphical model can be mapped to the physical network structure, thus enabling *decentralized implementation*.

1.2 APPLICATIONS

Some examples of inference problems in wireless networks are summarized in Table 1.1. The first example (signal detection) is a classic signal processing problem, with traditional application in radar systems, and of great interest nowadays for its application to spectrum sensing in CR networks. Since the variable to be estimated is binary, signal detection is formulated as a binary hypothesis testing problem. Localization is, on the contrary, an example of inference problem involving continuous variables: the position of a wireless device is a 2-D or 3-D vector. The third example can be thought of as an extension of signal detection to a multiple hypothesis setting, where the status of an object or event is modeled as a discrete variable taking values in a M-ary alphabet.

This work presented in this thesis addresses in particular the problem of signal detection, applied to CR networks, and the problem of cooperative localization combining peer-to-peer and GPS measurements (*hybrid* localization). A more detailed description of these two problems and their contexts of application is provided next.

1.2.1 *Spectrum Sensing in Cognitive Radio Networks*

The concept of dynamic spectrum allocation (DSA) was proposed as a part of the CR paradigm, originally introduced by J. Mitola in 1999 ([81]; see also [43, 1]). The DSA approach is based on the coexistence of different types of users in the same frequency bands: *primary* (licensed) users, that are the authorized owners of a given band, and *secondary* (unlicensed) users, that are allowed to opportunistically reuse the band when primary users are silent. The idea of DSA is motivated by extensive measurements performed by the Federal Communications Commission (FCC), revealing that fixed frequency assignment is highly inefficient both from a temporal and from a spatial point of view: in other words, some parts of the spectrum are strongly underutilized in certain areas and/or time of the day, whereas other frequencies (e.g., the 2.4 GHz band) are overcrowded. Therefore, DSA is seen as a promising solution to enable a more efficient spectrum usage, leading to higher data rates and more room for different, coexisting wireless systems.

In a DSA network, secondary users (SUs) must constantly monitor the spectrum, in order to identify unused spectrum portions (usually called *white spaces* or *spectrum holes*) and to promptly vacate them when a primary user (PU) starts transmission. This process, called **spectrum sensing**, is the key enabling feature of DSA and, by extension, of CR systems. Ideally, spectrum sensing should be efficient, reliable and reactive at the same time. Design of high-performance sensing techniques has been one of the most important research challenges of the past decade in the domain of wireless communications (see, for instance, [39, 121, 13, 168]).

The main challenge in spectrum sensing is the detection of low-power signals with high reliability. For example, the IEEE 802.22 WRAN standard [47], which regulates the use of white spaces in U.S. TV bands, requires detection of wireless microphone signals with power as low as -118dBm with error probabilities (both false-alarm and missed-detection) not exceeding 10%. Although such stringent sensing requirements were later removed by an FCC decision in 2010 [33] in favor of a database-oriented approach, spectrum sensing is still expected to play a major role in next-generation wireless technologies.

One way to improve the performance of spectrum sensing is collaboration among multiple devices: *collaborative* (or *cooperative*) spectrum sensing schemes, first introduced in [40, 80], provide increased robustness to fading or physical obstacles ("hidden node problem"), and substantially increase detection probability compared to single-user sensing. Following the definitions given in Sec. 1.1, cooperative sensing is a distributed inference problem (in particular, a problem of binary detection). This thesis addresses the problem of multi-sensor signal detection in various configurations:

- *Homogeneous, centralized* signal detection: all sensors observe the same signal and send the *received signal samples to a fusion center*; a covariance

matrix is constructed from all received signal samples. Note: sensors may be in this case multiple antennas of a single receiver.

- *Homogeneous, centralized* signal detection with *hard individual decisions* sent to the fusion center, with possibly *malicious users* (i.e., users that send wrong sensing reports in order to alter the final decision on spectrum occupancy). In this case, the goal is for the fusion center to maximize the detection performance while identifying and properly counteracting the attacks.

- *Heterogeneous, decentralized* signal detection: sensors in different locations may observe different signals, originated from different PUs. Thus, instead of a unique hidden variable indicating signal presence/absence for the entire network, here each sensor has to infer the value of a different binary state variable. The problem can be properly addressed by a decentralized algorithm, exploiting statistical correlations among neighboring nodes.

1.2.2 *Hybrid Cooperative Localization*

Global navigation satellite systems (GNSS) are the state-of-the art technology to perform localization in wireless networks, and their usage is widespread in a variety of civilian and military applications. In particular, the Global Positioning System (GPS) [54], developed in 1973 by the U.S. Department of Defense, has provided military support since the first Gulf war and is commonly used in vehicle navigation and portable devices.

However, global navigation satellite systems (GNSS) fail in some critical scenarios, such as in natural/urban canyons, under tree canopies, indoors, at high latitudes, in presence of electromagnetic interference. The failure, in the first three cases, is due to satellite signals being unable to sufficiently penetrate physical obstacles; in case of high latitude, it is due to geometric conditions (satellite orbits range from 0° to 55° latitude), so that the received GPS signal is affected by strong uncertainty; finally, harmful interference may result from terrestrial wireless systems or intentional jamming. The above shortcomings result in reduced localization availability and/or accuracy in several practical scenarios and, in military applications, can lead to a loss of lives on the ground as well as a tactical disadvantage in urban warfare [41].

To cope with this problem, odometers and inertial sensors can be used to track the receiver position via dead reckoning when GNSS is unavailable [116]. The main problem with dead reckoning is the error drift over time, so that the sensors are reliable for only very short periods of time (on the order of some seconds). Alternatively, the use of *hybrid positioning* has been proposed [79], where the receiver relies on terrestrial radio-frequency (RF) signals, such as GSM/UMTS, DVB, and Wi-Fi, in order to extract position-related measurements (e.g., signal strength, time of flight, or Doppler shift). Such

methods, called "systems of opportunity", rely on the presence of fixed terrestrial infrastructure, which may not always be available.

To reduce the reliance on fixed infrastructure, *peer-to-peer cooperative* methods have been proposed in [48, 19, 158], especially in the context of indoor wireless ad-hoc and sensor networks. Typically, GNSS data is not exploited by devices, thus these methods still require a limited local infrastructure to provide a geographical reference. In this thesis, the case of *hybrid* GNSS-cooperative localization is considered, where information received from satellites is combined with peer-to-peer range measurements, exchanged among network nodes. This approach, proposed in [101, 14, 16], requires devices to be equipped with a multifunction receiver comprising *(i)* a GNSS receiver; *(ii)* an RF system for short/medium distance peer-to-peer ranging (e.g., Ultra Wide-Band (UWB) or Wi-Fi); and *(iii)* a communication system (which may coincide with the ranging system). In this way, wireless networks achieve increased positioning availability for GNSS-denied nodes and increased accuracy for GNSS-challenged nodes.

Hybrid cooperative localization can be naturally formulated as a distributed *heterogeneous inference problem*, where hidden variables to be estimated are nodes' positions (2-D or 3-D vectors) along with nodes' biases with respect to satellite clock. It will be shown that in this case the statistical problem structure admits a one-to-one mapping to the physical network, therefore the problem can be addressed by a fully *decentralized architecture* without a fusion center.

1.3 Structure and Contributions of this Thesis

The thesis is organized as follows: Part ii (Chapters 2 through 4) is dedicated to homogeneous, multi-sensor signal detection for CR networks; Part iii (Chapters 5 through 9) focuses on belief propagation and its application to distributed and decentralized inference in wireless networks. In particular:

- Chapters 2 and 3 investigate detection methods based on the received signal covariance matrix, applying tools from RMT. The work presented in this part builds upon previous works, in particular [167, 17, 9, 73], where eigenvalue-based detection techniques were first introduced. Original contributions are performance analysis and improved design expressions of the aforementioned detectors. Specifically, results include:
  1. Analysis of false-alarm probability of the eigenvalue-ratio detector (ERD), and accurate expression of the decision threshold as a function of the target false-alarm rate.
  2. Performance analysis (detection probability) of the ERD.
  3. Performance analysis (detection probability) of the generalized likelihood ratio test (GLRT).

4. Comparison of GLRT and Roy's largest root test (RLRT) performance, and quantification of the impact of noise level knowledge.
5. Introduction of Hybrid Roy's largest root test (HRLRT), a detector able to close the gap between GLRT and RLRT by exploiting past observations.

- Chapter 4 presents, as a case study, an application of the CR paradigm to WSNs or WPANs operating in the 2.4 GHz ISM band. Contributions include theoretical results (eigenvalue-based and energy-based detection of discontinuous signals, design of *ad hoc* spectrum sensing strategies, optimization of sensing period) and implementation of a frequency-agile wireless sensor network (FA-WSN) prototype.

- Chapter 5 introduces probabilistic graphical models, the BP algorithm, and explains how BP can be applied to distributed inference in wireless networks allowing, sometimes, for fully decentralized implementations.

- Chapter 6 presents hybrid cooperative localization techniques based on the BP algorithm run over a wireless network. Contributions are:
  1. Development of a mathematical framework for hybrid cooperative localization, including a mapping between the network and a factor graph (FG).
  2. Derivation of a decentralized positioning algorithm, named hybrid sum-Product algorithm over a wireless network (H-SPAWN), based on BP on the previously defined FG model.
  3. Introduction of an efficient parametric message representation to reduce computational complexity and communication overhead.
  4. Performance comparison of the proposed algorithm with two algorithms based on conventional estimation methods: least squares estimation and Kalman filtering.
  5. Derivation of the Cramér-Rao bound (CRB), i.e., the lower bound of the achievable performance of any (unbiased) hybrid localization technique.

- Chapter 7 is focused on decentralized BP as a tool to perform distributed signal detection in heterogeneous radio environments, with scattered sensors possibly observing different signal sources. Contributions include:
  1. Development of a probabilistic graphical model based on Markov random fields (MRFs) for heterogeneous signal detection, and mapping of the MRF to the physical network.
  2. Derivation of a decentralized signal detection algorithm, that combines a peer-to-peer collaborative spectrum sensing approach (based

on BP over the previously defined MRF) with local Neyman-Pearson tests, allowing for a precise control of the false-alarm probability of each node.

3. Introduction of a learning method that results in automatic clustering of nodes which experience similar radio conditions (thus achieving performance similar to cluster-wise cooperative energy detection, without fusion centers and without prior knowledge of clusters).

- Chapter 8 applies graphical models (FGs and the BP algorithm) to the problem of cooperative signal detection (spectrum sensing) in the presence of misbehaving sensors. Contributions are:
  1. Definition of a statistical attack model for malicious sensors in cooperative spectrum sensing.
  2. Study of the impact of the considered attack model on the performance of cooperative spectrum sensing.
  3. Development of a Bayesian approach that combines the tasks of spectrum sensing, identification of malicious nodes, and estimation of attack probabilities of such nodes.
  4. Derivation of a BP algorithm that, adopted by the fusion center, performs efficiently the above Bayesian joint estimation. Two versions of the algorithm ("joint" and "sequential") are proposed.
  5. Theoretical and numerical performance analysis of the proposed algorithm. The key result is that, if nodes behave maliciously only with a certain probability, they should not be excluded from the decision process, but their data (properly weighted based on estimation of the attack probability) can be beneficial in terms of detection performance.

- Chapter 9 investigates "reweighted" variants of the classic BP algorithm, providing superior performance in many scenarios of practical interest. In particular, a novel technique, called uniformly reweighted belief propagation (URW-BP), is proposed. URW-BP lends itself well to decentralized implementation and is shown to outperform standard BP in applications such as distributed detection and cooperative localization when FGs contain cycles.

Finally, Part iv contains summarizing conclusions and remarks.

Part II

MULTI-SENSOR SIGNAL DETECTION AND RANDOM MATRIX THEORY

2 PROBLEM FORMULATION

2.1 NOTATION

In the following, upper-case boldface letters indicate matrices, lower-case boldface letters indicate vectors, symbols $^{\mathsf{T}}$ and $^{\mathsf{H}}$ indicate respectively the transpose and conjugate transpose (Hermitian) operators, $\mathrm{tr}(\cdot)$ is the trace of a matrix, $\|\cdot\|$ is the Euclidean norm of a vector, $\mathrm{diag}(\boldsymbol{x})$ indicates a square diagonal matrix whose main diagonal entries are taken from the vector $\boldsymbol{x}$, $\boldsymbol{I}_N$ is the identity matrix (of size N if specified), $\mathbf{0}_{M,N}$ and $\mathbf{1}_{M,N}$ are $M \times N$ matrices of zeros or ones, respectively; the symbol $\triangleq$ stands for "defined as", the symbol $\sim$ for "distributed with law", $\xrightarrow{\text{a.s.}}$ indicates almost sure convergence, and $\xrightarrow{D}$ convergence in distribution; $\mathrm{I}_{\{\alpha\}}$ is the indicator function which takes value 1 when condition α is true and 0 otherwise.

2.2 MATHEMATICAL MODEL

Consider a *cooperative detection* framework in which K receivers (or antennas) collaborate to sense the spectrum, which may be occupied by P primary signals (with $K > P$)[1].

Denote with y_k be the discrete baseband complex sample at receiver k, and define the $K \times 1$ received vector $\boldsymbol{y} = [y_1 \dots y_K]^{\mathsf{T}}$ containing the received signal samples at K sensors. The goal of the detector is to discriminate between two hypotheses:

- $\mathcal{H}_0$ (**absence of primary signal**). The received vector contains only noise:

$$\boldsymbol{y}|_{\mathcal{H}_0} = \boldsymbol{v} \tag{2.1}$$

- $\mathcal{H}_1$ (**presence of primary signal**). The received vector contains both signal and noise:

$$\boldsymbol{y}|_{\mathcal{H}_1} = \boldsymbol{x} + \boldsymbol{v} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{v} \tag{2.2}$$

  where:

1 As will be clarified in the following, the condition $K > P$ is required by covariance-based methods, because if $K \leqslant P$ the sample covariance matrix lacks the necessary degrees of freedom to distinguish signal components from noise. If the exact value of P is unknown, K must be chosen greater than the maximum possible number of primary signals.

- The $K \times 1$ vector $\boldsymbol{v} = [v_1 \dots v_K]^T \sim \mathcal{N}_C(\boldsymbol{0}_{K,1}, \sigma_v^2 \boldsymbol{I}_K)$ is a vector of circularly symmetric complex Gaussian (CSCG) noise samples, each with variance σ_v^2.
- The $P \times 1$ vector $\boldsymbol{s} = [s_1 \dots s_P]^T$ contains the primary signal samples originated from P different signal sources (PUs). The primary signals are assumed to be complex, zero-mean and mutually independent with covariance matrix

$$\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^H] = \operatorname{diag}(\sigma_1^2, \dots, \sigma_P^2) \triangleq \boldsymbol{\Sigma} \tag{2.3}$$

  where σ_p^2 is the variance of the p-th primary signal.
- The $K \times P$ complex matrix $\boldsymbol{H} \triangleq [\boldsymbol{h}_1 \dots \boldsymbol{h}_P]$ is the channel matrix, whose columns $\boldsymbol{h}_p = \left[h_{p1}, \dots, h_{pK}\right]^T$ are the channel vectors relative to primary user $p = 1, \dots, P$.

The signal-to-noise ratio (SNR) is defined (under $\mathcal{H}_1$) as

$$\rho \triangleq \frac{\mathbb{E}\,\|\boldsymbol{x}\|^2}{\mathbb{E}\,\|\boldsymbol{v}\|^2} = \frac{\mathbb{E}\,(|x_1|^2 + \dots + |x_K|^2)}{\mathbb{E}\,(|v_1|^2 + \dots + |v_K|^2)}. \tag{2.4}$$

By definition:

$$\rho = \frac{\operatorname{tr} \boldsymbol{R}_{\boldsymbol{x}}}{K\sigma_v^2} = \frac{\operatorname{tr} \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H}{K\sigma_v^2} = \frac{\sum_{p=1}^{P} \|\boldsymbol{h}_p\|^2 \sigma_p^2}{K\sigma_v^2} = \sum_{p=1}^{P} \rho_p \tag{2.5}$$

where the SNR of user p is defined as

$$\rho_p \triangleq \frac{\|\boldsymbol{h}_p\|^2 \sigma_p^2}{K\sigma_v^2} \tag{2.6}$$

2.2.1 *The Statistical Covariance Matrix and its Spectral Properties*

Consider the statistical covariance matrix of the signal vector $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{s}$:

$$\boldsymbol{R}_{\boldsymbol{x}} \triangleq \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^H] = \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H. \tag{2.7}$$

Let $\zeta_1 \geqslant \dots \geqslant \zeta_K$ be the eigenvalues of $\boldsymbol{R}_{\boldsymbol{x}}$ (without loss of generality, sorted in decreasing order), and $\boldsymbol{\zeta} \triangleq [\zeta_1, \dots, \zeta_K]$. Since the signal covariance matrix $\boldsymbol{R}_{\boldsymbol{x}}$ has rank P, $\boldsymbol{\zeta}$ contains $K - P$ zero eigenvalues: $\boldsymbol{\zeta} = [\zeta_1, \dots, \zeta_P, 0, \dots, 0]$. The remaining P non-zero eigenvalues ζ_i are expressed by the following property.

Lemma 1. *The P non-zero eigenvalues $\zeta_1, \dots, \zeta_P$ are the roots of the (reduced) degree P equation:*

$$\det\left(\boldsymbol{H}^H\boldsymbol{H} - z\,\boldsymbol{\Sigma}^{-1}\right) = 0 \tag{2.8}$$

where $\boldsymbol{\Sigma}^{-1} = \operatorname{diag}(\sigma_1^{-2}, \dots, \sigma_P^{-2})$.

Proof. The eigenvalues $\boldsymbol{\zeta}$ are the roots of the characteristic equation $\det(\boldsymbol{R_x} - z\boldsymbol{I}_K) = 0$ which can be simplified by applying the generalized matrix determinant lemma [12]:

$$\begin{aligned}
&\det\left(\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H - z\boldsymbol{I}_K\right) = \\
&= \det(\boldsymbol{\Sigma})\,\det(-z\boldsymbol{I}_K)\,\det\left(\boldsymbol{\Sigma}^{-1} - \frac{1}{z}\boldsymbol{H}^H\boldsymbol{H}\right) = \\
&= \left(\prod_{p=1}^{P}\sigma_p^2\right)(-z)^{K-P}\,\det\left(\boldsymbol{H}^H\boldsymbol{H} - z\,\boldsymbol{\Sigma}^{-1}\right)
\end{aligned} \tag{2.9}$$

where the first factor is constant with respect to z and the second term gives rise to the $(K-P)$ trivial solutions $\zeta_{P+1},\dots,\zeta_K = 0$. Since $\boldsymbol{\Sigma}$ is diagonal, $\boldsymbol{\Sigma}^{-1} = \mathrm{diag}(\sigma_1^{-2},\dots,\sigma_P^{-2})$. The non-zero signal eigenvalues $\zeta_1,\dots,\zeta_P$ thus result from the reduced-degree characteristic equation (2.8), which has degree P instead of K. □

Let us now consider the statistical covariance matrix of the received vector $\boldsymbol{y}$:

$$\boldsymbol{R} \triangleq \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^H]. \tag{2.10}$$

Because of (2.1) and (2.2), the covariance matrix is equal to, respectively:

$$\boldsymbol{R} = \begin{cases} \sigma_v^2\boldsymbol{I}_K & (\mathcal{H}_0) \\ \boldsymbol{R_x} + \sigma_v^2\boldsymbol{I}_K & (\mathcal{H}_1) \end{cases} \tag{2.11}$$

Let $\ell_1 \geqslant \dots \geqslant \ell_K$ be the eigenvalues of $\boldsymbol{R}$ (sorted in decreasing order), and $\boldsymbol{\ell} \triangleq [\ell_1,\dots,\ell_K]$ the corresponding vector. It can be immediately verified that under $\mathcal{H}_0$ all the eigenvalues of $\boldsymbol{R}$ are equal to the noise variance, while under $\mathcal{H}_1$ the first P eigenvalues contain a signal component:

$$\boldsymbol{\ell} = \begin{cases} \sigma_v^2\mathbf{1}_{1,K} = \left[\sigma_v^2,\dots,\sigma_v^2\right] & (\mathcal{H}_0) \\ \boldsymbol{\zeta} + \sigma_v^2\mathbf{1}_{1,K} = \left[\zeta_1+\sigma_v^2,\dots,\zeta_P+\sigma_v^2,\sigma_v^2,\dots,\sigma_v^2\right] & (\mathcal{H}_1) \end{cases} \tag{2.12}$$

The Single Signal Case ($P = 1$)

The single signal case is by far the most important scenario of interest in CR, because *(i)* typically a SU observes one PU signal only, and *(ii)* if multiple signal are present at the same time, detection performance depends essentially on the strongest one (i.e., the nearest PU).

When $P = 1$, a number of simplifications apply. First of all the channel matrix $\boldsymbol{H}$ is simply a vector $\boldsymbol{h}_1 = [h_{11}\dots h_{1K}]^T$ and the signal vector $\boldsymbol{s}$ is a scalar: $\boldsymbol{x} = \boldsymbol{h}_1 s_1$.

Thus, the signal covariance matrix reduces to

$$\boldsymbol{R_x} = \sigma_1^2\boldsymbol{h}_1\boldsymbol{h}_1^H \tag{2.13}$$

and the unique signal eigenvalue is readily found to be

$$\zeta_1|_{P=1} = \|\boldsymbol{h}_1\|^2\sigma_1^2 \tag{2.14}$$

Finally, the SNR simplifies to:

$$\rho|_{P=1} = \frac{\|\boldsymbol{h}_1\|^2\sigma_1^2}{K\sigma_v^2}. \tag{2.15}$$

2.2.2 *Sample Covariance Matrix*

In practice, the statistical correlation matrix $\boldsymbol{R}$ is estimated through a *sample covariance matrix*. Define the parameter N as the number of samples collected by each receiver during the sensing window. It is assumed that consecutive samples are uncorrelated and that, within the sensing window, all involved random processes (signals and noise) remain stationary and the channels are memoryless and constant.

Denote by $\boldsymbol{s}(n)$ the transmitted $P \times 1$ source vector at time n and define the $P \times N$ matrix

$$\boldsymbol{S} \triangleq [\boldsymbol{s}(1) \dots \boldsymbol{s}(N)] \tag{2.16}$$

Similarly, let $\boldsymbol{v}(n)$ and $\boldsymbol{y}(n)$ be, respectively, the noise and received vectors (of size $K \times 1$) at time n, and the corresponding $K \times N$ matrices

$$\boldsymbol{V} \triangleq [\boldsymbol{v}(1) \dots \boldsymbol{v}(N)] \tag{2.17}$$

$$\boldsymbol{Y} \triangleq [\boldsymbol{y}(1) \dots \boldsymbol{y}(N)] = \boldsymbol{H}\boldsymbol{S} + \boldsymbol{V} \tag{2.18}$$

The $K \times K$ sample covariance matrix $\boldsymbol{R}(N)$ is thus defined as

$$\boldsymbol{R}(N) \triangleq \frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^H \tag{2.19}$$

Denote with $\lambda_1 \geqslant \dots \geqslant \lambda_K$ the eigenvalues of $\boldsymbol{R}(N)$ (without loss of generality, sorted in decreasing order). *Eigenvalue-based detectors* use such eigenvalues to infer the presence of signal.

2.3 EIGEN-BASED DETECTION

Ideally, as $N \to \infty$, the sample covariance matrix $\boldsymbol{R}(N)$ converges to the statistical covariance matrix $\boldsymbol{R}$. The spectral properties of $\boldsymbol{R}$, summarized by (2.12), motivate the adoption of the *largest sample eigenvalue* (λ_1) as a test statistic to discriminate between hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$: it would be sufficient to check whether λ_1 is equal to the other eigenvalues. For finite N, on the contrary, sample eigenvalues are random variables characterized by a probabilistic behavior, which gives rise to possible detection errors.

The following tests have been proposed in the CR literature:

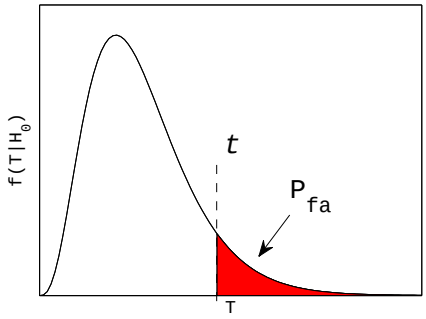


Figure 2.1: False-alarm probability for a generic test statistic T.

- **Roy's largest root test** (RLRT), originally developed in [120], and applied to spectrum sensing in [150]:

$$T_{\text{RLRT}} \triangleq \frac{\lambda_1}{\sigma_v^2}. \tag{2.20}$$

- **Eigenvalue-ratio detector** (ERD), introduced in the CR literature by [167, 17, 166]:

$$T_{\text{ERD}} \triangleq \frac{\lambda_1}{\lambda_K}. \tag{2.21}$$

- **Generalized likelihood ratio test** (GLRT), introduced quite recently in signal processing [7] and CR [9, 73], but previously proposed and studied in the statistics literature, e.g., [52, 126]:

$$T_{\text{GLRT}} \triangleq \frac{\lambda_1}{\text{tr}\,(\boldsymbol{R}(N))}. \tag{2.22}$$

  Remark. The term GLRT is very general and can be applied to a variety of estimation problems. In this thesis, GLRT is referred to the specific detector proposed in [9], and derived from the signal model introduced in Sec. 2.2 assuming single signal ($P = 1$), unknown channel vector $\boldsymbol{h}$, and unknown noise variance σ_v^2. Other GLRT detectors have been later introduced in [130] under different assumptions.

Following the terminology of [168], the RLRT is "semi-blind", as it requires exact knowledge of the noise variance, whereas the other tests are "blind", as they implicitly estimate the noise variance from the data.

Given a detector that uses a generic test statistic T and a decision threshold t, performance can be characterized by two probabilities.

- *Probability of false alarm* (signal is absent but it is erroneously detected), defined as

$$P_{\text{fa}} \triangleq \Pr(T > t|\mathcal{H}_0) = 1 - F_{T|\mathcal{H}_0}(t). \tag{2.23}$$

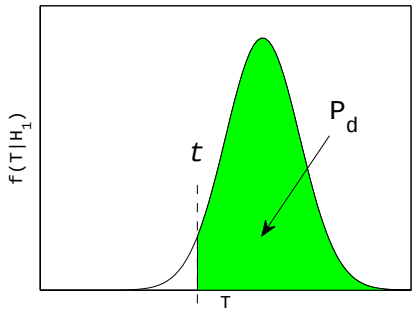


Figure 2.2: Detection probability for a generic test statistic T.

- *Probability of detection* (signal is present and correctly detected), defined as

$$P_d \triangleq \Pr(T > t|\mathcal{H}_1) = F_{T|\mathcal{H}_1}(t). \tag{2.24}$$

The symbol $F_{T|\mathcal{H}_i}(t)$ in (2.23) and (2.24) denotes the cumulative distribution function (CDF) of T under hypothesis $\mathcal{H}_i$. The *probability of missed detection* can be then defined as the complementary of P_d:

$$P_{md} \triangleq 1 - P_d. \tag{2.25}$$

Detection and false-alarm probability are represented graphicaly in Figs. 2.1 and 2.2. In order to evaluate the performance of the considered detectors, it is necessary to express analytically the distributions of the corresponding test statistics T_{RLRT}, T_{ERD}, and T_{GLRT}, under both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. This is done in the next chapter by applying asymptotic results from RMT.

3

PERFORMANCE ANALYSIS OF EIGEN-BASED DETECTORS

In this chapter, the performance of the eigenvalue-based detectors introduced in Sec. 2.3 is analyzed using RMT tools. Preliminary mathematical result are first presented in Sec. 3.1. Then, false-alarm and detection probabilities of RLRT, ERD, and GLRT are expressed in Sec. 3.2. Finally, the impact of noise level knowledge is studied in Sec. 3.3.

3.1 RESULTS FROM RANDOM MATRIX THEORY

3.1.1 *Definitions*

The following distribution functions will be used in the following.

3.1.1.1 *Tracy-Widom Distribution*

The Tracy-Widom distributions were first introduced in [136]. The second-order Tracy-Widom CDF, $F_{\mathcal{W}_2}(x)$, can be defined as

$$F_{\mathcal{W}_2}(x) = \det(1 - \boldsymbol{A}_x)$$

where $\boldsymbol{A}_x$ is the operator acting on $L^2((x, +\infty))$ with Airy kernel $A(u, v) = \frac{Ai(u)Ai'(v) - Ai'(u)Ai(v)}{u - v}$ where $Ai(u) = \frac{1}{2\pi} \int_{\infty e^{5j\pi/6}}^{\infty e^{j\pi/6}} e^{jua + j\frac{1}{3}a^3} da$ is the complex Airy function.

The Tracy-Widom distribution also admits an alternative expression. Let $q(u)$ be the solution of the Painlevé II differential equation

$$q''(u) = uq(u) + 2q^3(u) \tag{3.1}$$

satisfying

$$q(u) \sim -Ai(u), \quad u \to +\infty \tag{3.2}$$

Then

$$F_{\mathcal{W}_2}(x) = \exp\left(-\int_x^{+\infty} (u - x) q^2(u) du\right) \tag{3.3}$$

3.1.1.2 *Airy-type Distributions*

These distributions are defined in [5] as an extension of the Tracy-Widom distribution for $k \geqslant 1$

$$F_{\mathcal{A}_k}(x) = \det(1 - \boldsymbol{A}_x) \cdot \\ \cdot \det\left(\delta_{mn} - < \frac{1}{1-\boldsymbol{A}_x} s^{(m)}, t^{(n)} >\right)_{1 \leqslant m,n \leqslant k}$$

where $<,>$ is the real inner product of functions in $L^2((x, +\infty))$, and

$$s^{(m)}(u) = \frac{1}{2\pi} \int_{\infty e^{5j\pi/6}}^{\infty e^{j\pi/6}} e^{jua + j\frac{1}{3}a^3} \frac{1}{(ja)^m} da$$

$$t^{(m)}(u) = \frac{1}{2\pi} \int_{\infty e^{5j\pi/6}}^{\infty e^{j\pi/6}} e^{jua + j\frac{1}{3}a^3} (ja)^{m-1} da$$

For $k = 0$ it reduces to the $\mathcal{W}_2$.

3.1.1.3 *Finite Gaussian Unitary Ensemble Distributions*

$$F_{\mathcal{G}_k}(x) = (2\pi)^{-k/2} \left(\prod_{m=1}^{k} m!\right)^{-1} \cdot \\ \cdot \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \prod_{1 \leqslant m < n \leqslant k} |\xi_m - \xi_n|^2 \cdot \prod_{m=1}^{k} e^{-\frac{1}{2}\xi_m^2} d\xi_1 \ldots d\xi_k$$

In the case $k = 1$, it is simply a zero-mean, unit-variance Gaussian distribution:

$$F_{\mathcal{G}_1}(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-\frac{1}{2}\xi^2} d\xi$$

3.1.2 *Asymptotic eigenvalue distributions under $\mathcal{H}_0$*

Under $\mathcal{H}_0$, since the columns of $\boldsymbol{Y}$ are zero-mean independent complex Gaussian vectors, the sample covariance matrix $\boldsymbol{R}(N)$ is a *complex Wishart matrix* [155].

The fluctuations of the eigenvalues of Wishart matrices have been thoroughly investigated by RMT (see [137] and [3] for an overview). The key idea of RMT is that in many cases the eigenvalues of matrices with random entries turn out to converge to some fixed distribution, when both the dimensions of the signal matrix tend to infinity with the same order. For Wishart matrices the limiting joint eigenvalue distribution has been known for many years [75]; then, more recently, also the marginal distributions of single ordered eigenvalues have been found.

By exploiting some of these results, it is possible to express the asymptotical values of the largest and the smallest eigenvalue of $\boldsymbol{R}(N)$ as well as their limiting distributions. The following theorem summarizes a number of relevant results.

Theorem 1. ***Convergence of the smallest and largest eigenvalues under*** $\mathcal{H}_0$**.** *Let*

$$c \triangleq \frac{K}{N} \tag{3.4}$$

and assume that for $K, N \to \infty$

$$c \to \bar{c} \in (0, 1) \tag{3.5}$$

Define:

$$\mu_{\pm}(c) \triangleq \left(c^{1/2} \pm 1\right)^2 \tag{3.6}$$

$$\nu_{\pm}(c) \triangleq \left(c^{1/2} \pm 1\right)\left(c^{-1/2} \pm 1\right)^{1/3} \tag{3.7}$$

Then, as $N, K \to \infty$*, the following holds:*

(i) Almost sure convergence of the largest eigenvalue

$$\lambda_1 \xrightarrow{a.s.} \sigma_v^2\, \mu_+(c) \tag{3.8}$$

(ii) Convergence in distribution of the largest eigenvalue

$$N^{2/3}\, \frac{\lambda_1 - \sigma_v^2\, \mu_+(c)}{\sigma_v^2\, \nu_+(c)} \xrightarrow{D} \mathcal{W}_2 \tag{3.9}$$

(iii) Almost sure convergence of the smallest eigenvalue

$$\lambda_K \xrightarrow{a.s.} \sigma_v^2\, \mu_-(c) \tag{3.10}$$

(iv) Convergence in distribution of the smallest eigenvalue

$$N^{2/3}\, \frac{\lambda_K - \sigma_v^2\, \mu_-(c)}{\sigma_v^2\, \nu_-(c)} \xrightarrow{D} \mathcal{W}_2 \tag{3.11}$$

where $\mathcal{W}_2$ *is the Tracy-Widom law of order 2, defined in Appendix 3.1.1.1.*

Proof. The claims of this theorem follow from different results of RMT, up to some changes of variables and using a uniform notation. Proofs of the original theorems appear in the references listed below.
Claims (i) and (iii) descend from the work by Marchenko and Pastur [75], later extended by Silverstein, Bai, Yin, *et al.* [3].
Claim (ii) was proved, under the assumption of Gaussian entries, by Johansson [51], Johnstone [53] and Soshnikov [129], and generalized to the non-Gaussian case by Péché [96].
Claim (iv) derives from a very recent result by Feldheim and Sodin [34]. □

3.1.3 *Case $\mathcal{H}_1$: Reduction to Spiked Population Model*

Under $\mathcal{H}_1$, the received signal matrix $\boldsymbol{Y}$ contains some Gaussian entries, like in the Wishart case, along with a certain number (P) of signal components. In order to put into evidence the spiked structure of $\boldsymbol{R}(N)$, the received signal matrix $\boldsymbol{Y}$ (2.17) needs to be rewritten in the form

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{Z} \tag{3.12}$$

where $\boldsymbol{T}$ is a block matrix of size $K \times (P + K)$ defined as

$$\boldsymbol{T} = \left[\begin{array}{c|c} \frac{1}{\sigma_v}\boldsymbol{H}\boldsymbol{\Sigma}^{1/2} & \boldsymbol{I}_K \end{array} \right] \tag{3.13}$$

and $\boldsymbol{Z}$, of size $(P + K) \times N$, is defined as

$$\boldsymbol{Z} = \left[\begin{array}{c} \sigma_v \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S} \\ \hline \boldsymbol{V} \end{array} \right] \tag{3.14}$$

This decomposition has been chosen such that all the entries z_{ij} of $\boldsymbol{Z}$ ($1 \leqslant i \leqslant P + K$, $1 \leqslant j \leqslant N$) have the following properties:

$$\mathbb{E}\, z_{ij} = 0 \tag{3.15}$$

$$\mathbb{E}\, |z_{ij}|^2 = \sigma_v^2 \tag{3.16}$$

which are necessary conditions for Theorem 2 to hold. The covariance matrix becomes then

$$\boldsymbol{R}(N) = \frac{1}{N}\boldsymbol{T}\boldsymbol{Z}\boldsymbol{Z}^H\boldsymbol{T}^H \tag{3.17}$$

which is exactly the model of [6], [5] and [35].

Finally, denote with $t_1, \ldots, t_K$ the eigenvalues of $\boldsymbol{T}\boldsymbol{T}^H$. It follows from the structure of $\boldsymbol{T}$ that P eigenvalues are different from 1 (without loss of generality, they are assigned the first P positions: $t_1 \geqslant \ldots \geqslant t_P$) and the remaining $K - P$ are ones. To express the P "spike eigenvalues" (that represent the perturbation with respect to the pure-noise model), it is observed that

$$\boldsymbol{T}\boldsymbol{T}^H = \frac{1}{\sigma_v^2}\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H + \boldsymbol{I}_K \tag{3.18}$$

and the eigenvalues $t_1, \ldots, t_P$ result from the solution of

$$\begin{aligned} &\det\left(\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H - \sigma_v^2(t-1)\boldsymbol{I}_K\right) = 0 \\ &\text{s.t. } t \neq 1 \end{aligned} \tag{3.19}$$

The structure of the problem is identical to that of (2.8), with the change of variable $z = \sigma_v^2(t-1)$. Therefore, the "spike eigenvalues" t_p are linked to the non-zero eigenvalues of the signal covariance matrix, ζ_p, by the relation

$$t_p = \frac{\zeta_p}{\sigma_v^2} + 1, \quad 1 \leqslant p \leqslant P \tag{3.20}$$

In general, the values of ζ_p are calculated using (2.9); in the case of *single primary user* ($P = 1$), the simplified expression (2.14) can be used, which yields

$$t_1|_{P=1} = \|\boldsymbol{h}_1\|^2 \frac{\sigma_1^2}{\sigma_\nu^2} + 1 \tag{3.21}$$

3.1.3.1 *Relation between spike eigenvalues and SNR*

The spike eigenvalues are related with the SNR; this fact turns out to be useful especially in the case of $P = 1$. From (3.20) it follows that

$$\sum_{p=1}^{P} t_p = \frac{1}{\sigma_\nu^2} \sum_{p=1}^{P} \zeta_p + P \tag{3.22}$$

but, from the eigendecomposition of $\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H$ and from (2.5) it follows that

$$\sum_{p=1}^{P} \zeta_p = \operatorname{tr} \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H = \rho K \sigma_\nu^2 \tag{3.23}$$

hence

$$\sum_{p=1}^{P} t_p = K\rho + P \tag{3.24}$$

Therefore, in the *case of one primary user* ($P = 1$), the (unique) spike eigenvalue may be expressed directly as a function of the SNR:

$$t_1|_{P=1} = K\rho + 1 \tag{3.25}$$

Note that, by exploiting the property (3.23), one could also obtain (2.14) without resorting to the characteristic equation.

In the *case of multiple primary signals* ($P > 1$), the *sum* of the spike eigenvalues is related to the SNR, but not the *single* eigenvalues. Therefore, to compute the t_p (in particular t_1, which is needed to apply Theorem 2), it is necessary to know the channel matrix and the power of primary signals and use (2.8).

3.1.4 *Asymptotic eigenvalue distributions under $\mathcal{H}_1$*

The following theorem provides a useful result on the convergence of the largest eigenvalue in spiked population models.

Theorem 2. ***Convergence of the largest eigenvalue under $\mathcal{H}_1$.*** *Again, assume that for* $K, N \to \infty$

$$c = \frac{K}{N} \to \bar{c} \in (0, 1) \tag{3.26}$$

In addition, assume that for all i, j *s.t.* $1 \leqslant i \leqslant P + K, \quad 1 \leqslant j \leqslant N$*:*

(A_1) $\mathbb{E}\, z_{ij} = 0$

(A_2) $\mathbb{E}\,(\Re\, z_{ij})^2 = \mathbb{E}\,(\Im\, z_{ij})^2 = \frac{\sigma_v^2}{2}$

(A_3) $\forall k > 0,\ \ \mathbb{E}\,|z_{ij}|^{2k} < \infty$ *and* $\mathbb{E}\,(\Re\, z_{ij})^{2k+1} = \mathbb{E}\,(\Im\, z_{ij})^{2k+1} = 0$

(A_4) $\mathbb{E}\,(\Re\, z_{ij})^4 = \mathbb{E}\,(\Im\, z_{ij})^4 = \frac{3}{4}\sigma_v^4$

Define:

$$\mu_s(t_1, c) \triangleq t_1 \left(1 + \frac{c}{t_1 - 1}\right) \tag{3.27}$$

$$\nu_s(t_1, c) \triangleq t_1 \sqrt{1 - \frac{c}{(t_1-1)^2}} \tag{3.28}$$

Then, as $N, K \to \infty$*, the following holds:*

(i) Almost sure convergence of the largest eigenvalue: phase transition phenomenon
If $t_1 > 1 + c^{1/2}$*:*

$$\lambda_1 \xrightarrow{a.s.} \sigma_v^2 \mu_s(t_1, c) \tag{3.29}$$

If $t_1 \leqslant 1 + c^{1/2}$*:*

$$\lambda_1 \xrightarrow{a.s.} \sigma_v^2\, \mu_+(c) \tag{3.30}$$

(ii) Convergence in distribution of the largest eigenvalue
Let m *(with* $1 \leqslant m \leqslant P$*) be the multiplicity of the first spike eigenvalue* t_1*.*
If $t_1 = \ldots = t_m > 1 + c^{1/2}$*:*

$$N^{1/2}\, \frac{\lambda_1 - \sigma_v^2\, \mu_s(t_1, c)}{\sigma_v^2\, \nu_s(t_1, c)} \xrightarrow{D} \mathcal{G}_m \tag{3.31}$$

If $t_1 = \ldots = t_m = 1 + c^{1/2}$*:*

$$N^{2/3}\, \frac{\lambda_1 - \sigma_v^2\, \mu_+(c)}{\sigma_v^2\, \nu_+(c)} \xrightarrow{D} \mathcal{A}_m \tag{3.32}$$

If $t_1 < 1 + c^{1/2}$*:*

$$N^{2/3}\, \frac{\lambda_1 - \sigma_v^2\, \mu_+(c)}{\sigma_v^2\, \nu_+(c)} \xrightarrow{D} \mathcal{W}_2 \tag{3.33}$$

where $\mathcal{A}_m$ *and* G_m *are distribution laws defined in Appendix 3.1.1.2 and 3.1.1.3, respectively.*

Proof. The proof of claim (i) is due to Baik and Silverstein [6]; claim (ii) was found by Baik, Ben Arous and Péché [5] under the additional assumption of z_{ij} Gaussian with unit variance, and was generalized into this form by Féral, Péché [35] using results from Bai and Yao [4]. □

3.1.4.1 *Remarks*

VALIDITY OF THE ASSUMPTIONS All the assumptions (A_1)-(A_4) are verified exactly for the noise part of $\boldsymbol{Z}$, whose entries are complex Gaussian random variables. For the signal part, the first two assumptions are guaranteed by construction of $\boldsymbol{Z}$: (A_1) is given by (3.15) and (A_2) is equivalent to (3.16) (provided that the variance of $\boldsymbol{s}$ is equally distributed between real and imaginary part, which is true for all types of complex signals used in communications). Assumption (A_3) is also verified in practical cases.

Assumption (A_4) is satisfied exactly by Gaussian signals, while there exist several types of signals (e.g. PSK, QAM) whose fourth moment is lower than that of a Gaussian random variable. However, since the type of primary signal is usually unknown at the secondary users, the Gaussian assumption is reasonable in general. In addition, since $P < K$, most of matrix $\boldsymbol{Z}$ is represented by the noise part which does always satisfy (A_4): therefore the theorem can be applied in almost all practical cases, even when this assumption does not hold exactly. The approximation introduced in this way is small and becomes negligible when the SNR of the primary signal is low, as shown by numerical simulations.

PHASE TRANSITION PHENOMENON The first important result implied by the theorem is the existence of a *critical value* of t_1 that determines whether a signal component is identifiable or not. This behavior is called *phase transition phenomenon*. In fact, when $t_1 \leqslant 1 + c^{1/2}$, the largest eigenvalue of the covariance matrix converges to the same value as in the pure-noise model, whereas for $t_1 > 1 + c^{1/2}$, it converges to a larger value: $\mu_s(t_1, c) > \mu_+(c)$. This property makes it possible to detect the presence of signals.

In case of $P = 1$, the critical value can be expressed directly in terms of the SNR using (3.25):

$$\rho > \frac{1}{\sqrt{KN}} \tag{3.34}$$

This relation also allows to determine the *minimum number of samples* for the detector to be able to identify signals with a given SNR.

LIMITING DISTRIBUTIONS The second claim of the theorem clarifies *how* the largest eigenvalue converges to the asymptotical limit. For non-identifiable components, the limiting distribution is the same as in the case of no signal. For components with eigenvalues placed exactly on the critical point, the limiting distribution is a generalization of the one encountered in the previous case: in fact, for $m = 0$, $\mathcal{A}_0$ reduces to the Tracy-Widom law (Sec.3.1.1.2). Components above the critical value result in distributions $\mathcal{G}_m$, which for $m = 1$ reduce to normal distributions. Note that both the events of eigenvalues exactly equal to the critical point and of eigenvalues with multiplicity larger than one are asymptotically *events with zero probability*. The results con-

cerning these cases are mentioned for completeness, but are not important for practical applications.

The following theorem gives insight into the asymptotic distribution of smallest eigenvalue.

Theorem 3. ***Distribution of the*** $K-P$ ***smallest eigenvalues under*** $\mathcal{H}_1$**.** *Assume that for* $K, N \to \infty$

$$c = \frac{K}{N} \to \bar{c} \in (0,1) \tag{3.35}$$

and that $t_p > 1 + c^{1/2}$ *for* $1 \leqslant p \leqslant P$*, the eigenvalues* $\lambda_{P+1}, \ldots, \lambda_K$ *of* $\boldsymbol{R}(N)$ *have asymptotically the same limiting distribution as those of a* $(K-P) \times (K-P)$ *Wishart matrix.*

Proof. The result follows from the proof of Lemma 2 in [60]. □

3.2 FALSE-ALARM AND DETECTION PROBABILITY ANALYSIS

3.2.1 *RLRT*

3.2.1.1 *False-alarm Probability and Decision Threshold*

The RMT results given in Sec. 3.1 make it possible to express directly the performance of the RLRT detector, in the asymptotic regime $K, N \to \infty$. Simulation results show that asymptotic expressions can be used as accurate approximations even for realistic numbers of samples N and sensors K.

From Theorem 1, the false-alarm probability of RLRT is given by

$$P_{fa}(\text{RLRT}) = \Pr\left(T_{\text{RLRT}} > t_R | \mathcal{H}_0\right) \approx 1 - F_{\mathcal{W}2}\left(\frac{t_R - \mu_+}{\xi_+}\right), \tag{3.36}$$

where t_R is the decision threshold, $F_{\mathcal{W}2}(\cdot)$ is the CDF of the second order Tracy-Widom distribution $\mathcal{W}_2$, μ_+ is given by (3.6), and

$$\xi_+ \triangleq N^{-2/3} \nu_+ \tag{3.37}$$

with ν_+ given by (3.7).

In practice, the decision threshold t_R can be selected as a function of a required false-alarm rate $P_{fa} = \alpha$, by inverting (3.36):

$$t_R(\alpha) = \mu_+ + \xi_+ F^{-1}_{\mathcal{W}2}(1-\alpha). \tag{3.38}$$

The values of $F^{-1}_{\mathcal{W}2}(\cdot)$ can be computed numerically (see for example [26]) and stored in a look-up table.

3.2.1.2 *Detection Probability*

Similarly, the RLRT detection probability is given asymptotically by Theorem 2. Assuming identifiable signals ($\rho > \frac{1}{\sqrt{KN}}$) and distinct eigenvalues (event with probability one for any number of signals P), the limiting distribution of λ_1 is a normal distribution, hence

$$P_d(\text{RLRT}) = \Pr\left(T_{\text{RLRT}} > t_R | \mathcal{H}_1\right) \approx Q\left(\sqrt{N}\frac{t_R - \mu_s}{\nu_s}\right). \tag{3.39}$$

where μ_s, ν_s are given by (3.27), (3.28), and where $Q(z) \triangleq \frac{1}{\sqrt{2\pi}}\int_z^\infty e^{-x^2/2}dx$ is the tail probability of a standard Gaussian distribution $\mathcal{N}(0,1)$.

In case of a single signal with SNR ρ, the above result can be written as

$$P_d(\text{RLRT})|_{P=1} \approx Q\left[\sqrt{N}\frac{t_R - (1+K\rho)\left(1+\frac{1}{N\rho}\right)}{(1+K\rho)\sqrt{1-\frac{1}{NK\rho^2}}}\right] \tag{3.40}$$

$$\approx Q\left[\sqrt{N}\left(\frac{t_R}{1+K\rho} - 1\right)\right]. \tag{3.41}$$

3.2.2 *ERD*

3.2.2.1 *False-alarm Probability and Decision Threshold*

The results of Theorem 1 allow, through some algebraic manipulations, to determine the limiting distribution of the test statistic T_{ERD} under $\mathcal{H}_0$.

In order to apply claims (ii) and (iv), define:

$$L_1 \triangleq N^{2/3}\,\frac{\lambda_1 - \sigma_v^2\,\mu_+(c)}{\sigma_v^2\,\nu_+(c)} \tag{3.42}$$

$$L_K \triangleq N^{2/3}\,\frac{\lambda_K - \sigma_v^2\,\mu_-(c)}{\sigma_v^2\,\nu_-(c)} \tag{3.43}$$

Both L_1 and L_K converge in distribution to the Tracy-Widom law $\mathcal{W}_2$:

$$f_{L_1}(z), f_{L_K}(z) \to f_{\mathcal{W}_2}(z) \tag{3.44}$$

where $f_{\mathcal{W}_2}(\cdot)$ represents the probability distribution function (PDF) associated with the law $\mathcal{W}_2$.

Then, T_{ERD} can be written as

$$T_{\text{ERD}} = \frac{\lambda_1}{\lambda_K} = \frac{N^{-2/3}\nu_+(c)L_1 + \mu_+(c)}{N^{-2/3}\nu_-(c)L_K + \mu_-(c)} \tag{3.45}$$

Notice that the term σ_v^2 is canceled out in the ratio (this is the reason that makes the ERD "blind" with respect to the noise power). Denote by l_1 and l_K , respectively, the numerator and the denominator of (3.45), and by $\bar{f}_{l_1}(z)$ and $\bar{f}_{l_K}(z)$ their limiting PDFs for $N, K \to \infty$. These distributions are the same as those of L_1 and L_K, up to a linear random variable transformation:

$$\bar{f}_{l_1}(z) = \frac{N^{2/3}}{\nu_+(c)} f_{\mathcal{W}_2}\left(\frac{N^{2/3}}{\nu_+(c)}(z - \mu_+(c))\right) \tag{3.46}$$

For the denominator, it must be observed that $\nu_-(c) < 0$ for the considered range $c \in (0,1)$. Thus

$$\begin{aligned}\bar{f}_{l_K}(z) &= \frac{N^{2/3}}{|\nu_-(c)|} f_{\mathcal{W}_2}\left(\frac{N^{2/3}}{|\nu_-(c)|}(\mu_-(c)-z)\right) \\ &= -\frac{N^{2/3}}{\nu_-(c)} f_{\mathcal{W}_2}\left(\frac{N^{2/3}}{\nu_-(c)}(z-\mu_-(c))\right)\end{aligned} \tag{3.47}$$

To express the distribution of T_{ERD}, observe that $f_{l_1}(l_1)$ and $f_{l_K}(l_K)$ are asymptotically independent in virtue of a classic result by Lawley [64] stating that the pairwise correlations of sample eigenvalues decreases at a rate $O(1/N)$. Thus,

$$\bar{f}_{l_1,l_K}(l_1,l_K) \approx \bar{f}_{l_1}(l_1)\bar{f}_{l_K}(l_K) \tag{3.48}$$

Then, using the formula for the quotient of random variables [24], the resulting ratio distribution writes:

$$\begin{aligned}\bar{f}_{T_{ERD}|\mathcal{H}_0}(t) &= \left[\int_{-\infty}^{+\infty} |x| \bar{f}_{l_1,l_K}(tx,x)dx\right] \cdot I_{\{t>1\}} \\ &= \left[\int_{0}^{+\infty} x \bar{f}_{l_1}(tx)\bar{f}_{l_K}(x)dx\right] \cdot I_{\{t>1\}}\end{aligned} \tag{3.49}$$

where the lower integration limit has been changed to 0 instead of $-\infty$, since the covariance matrix is positive-semidefinite therefore all the eigenvalues are non-negative; the condition $t > 1$ is necessary to preserve the order of the eigenvalues, since the distributions are defined under the assumption $l_1 > l_K$.

Finally, let $\bar{F}_{T_{ERD}|\mathcal{H}_0}(t)$ be the CDF corresponding to (3.49). For N and K large enough, the following approximation holds:

$$P_{fa}(ERD) = Pr(T_{ERD} > t_E|\mathcal{H}_0) \approx \bar{F}_{T_{ERD}|\mathcal{H}_0}(t_E). \tag{3.50}$$

The above expression can be inverted numerically to find the decision threshold t_E as a function of a false-alarm rate $P_{fa} = \alpha$. Similar as in the case of RLRT, the values of t_E of interest can be computed offline and stored in a look-up table.

Note that the distribution of T_{ERD} for finite N and K can also be expressed *exactly*, by adopting a different approach [111]. The drawback of the exact approach is its complexity, which makes implementation difficult when K and N are large.

3.2.2.2 *Detection Probability*

The detection probability of the ERD can be expressed by combining the results of Theorem 2 and of Theorem 3.

A similar approach as in the case of $\mathcal{H}_0$ is adopted. Define again

$$L_1 \triangleq N^{1/2} \frac{\lambda_1 - \sigma_v^2 \mu_s(t_1, c)}{\sigma_v^2 \nu_s(t_1, c)}. \tag{3.51}$$

By Theorem 2, assuming identifiable signals and distinct eigenvalues, variable L_1 has a limiting Guassian distribution:

$$f_{L_1}(z) \to \mathcal{N}(0, 1). \tag{3.52}$$

By Theorem 3, the distribution of the smallest eigenvalue is not affected by the presence of "spikes" and claims (iii) and (iv) of Theorem 1 can be applied also in this case with the only difference that, instead of c (3.4), now

$$c' = \frac{K - P}{N} \tag{3.53}$$

Thus,

$$L_K \triangleq N^{2/3} \frac{\lambda_K - \sigma_v^2 \mu_-(c')}{\sigma_v^2 \nu_-(c')} \tag{3.54}$$

still converges in distribution to the Tracy-Widom law

$$f_{L_K}(z) \to f_{\mathcal{W}_2}(z). \tag{3.55}$$

Then the test statistic T_{ERD} becomes

$$T_{ERD} = \frac{\lambda_1}{\lambda_K} = \frac{N^{-1/2}\nu_s(t_1, c)L_1 + \mu_s(t_1, c)}{N^{-2/3}\nu_-(c')L_K + \mu_-(c')} \tag{3.56}$$

Denote with l_1 and l_K, respectively, the numerator and the denominator of T_{ERD} and with $\bar{f}_{l_1}(z)$ and $\bar{f}_{l_K}(z)$ their limiting PDFs for $N, K \to \infty$. Through a random variable transformation, these distributions may be expressed as

$$\bar{f}_{l_1}(z) = \frac{(N/2\pi)^{1/2}}{\nu_s(t_1, c)} \exp\left[-\frac{N}{2\nu_s^2(t_1, c)}(z - \mu_s(t_1, c))^2\right] \tag{3.57}$$

$$\bar{f}_{l_K}(z) = \frac{N^{2/3}}{|\nu_-(c')|} f_{\mathcal{W}_2}\left(\frac{N^{2/3}}{|\nu_-(c')|}(\mu_-(c') - z)\right) \tag{3.58}$$

Since $f_{l_1}(l_1)$ and $f_{l_K}(l_K)$ are asymptotically independent, the resulting limiting ratio distributions is

$$\begin{aligned} \bar{f}_{T|\mathcal{H}_1}(t) &= \left[\int_{-\infty}^{+\infty} |x| \bar{f}_{l_1, l_K}(tx, x) dx\right] \cdot I_{\{t>1\}} \\ &= \left[\int_0^{+\infty} x \bar{f}_{l_1}(tx) \bar{f}_{l_K}(x) dx\right] \cdot I_{\{t>1\}} \end{aligned} \tag{3.59}$$

where, like in the previous case, the domain of integration has been restricted to non-negative values, and the condition $t > 1$ is necessary to ensure that $l_1 > l_K$.

Finally, denoting with $\bar{F}_{T|\mathcal{H}_1}(t)$ the CDF corresponding to the PDF in (3.59), the following approximation holds for K, N large enough:

$$P_d(ERD) = Pr(T_{ERD} > t_E | \mathcal{H}_1) \approx \bar{F}_{T_{ERD}|\mathcal{H}_1}(t_E), \tag{3.60}$$

where t_E is the ERD decision threshold.

3.2.3 *GLRT*

3.2.3.1 *False-alarm Probability and Decision Threshold*

Asymptotically, as both $N, K \to \infty$, the random variable T_{GLRT} follows a second-order Tracy-Widom distribution like T_{RLRT} [10], hence in first approximation the GLRT decision threshold $t_G(\alpha)$ can be computed like $t_R(\alpha)$ (3.38). However, as described in [87], this approximation is not very accurate for tail probabilities of T_{GLRT} for small values of K. In [87] the following improved expression was derived:

$$\Pr\left[\frac{T_{GLRT}-\mu_+}{\xi_+} < s\right] \approx F_{\mathcal{W}2}(s) - \frac{1}{2NK}\left(\frac{\mu_+}{\xi_+}\right)^2 F''_{\mathcal{W}2}(s),$$

where μ_+ is given in (3.6) and ξ_+ in (3.37). Hence,

$$\begin{aligned} P_{fa} = \alpha &= \Pr(T_{GLRT} > t_G) = \Pr\left(\frac{T_{GLRT}-\mu_+}{\xi_+} > \frac{t_G-\mu_+}{\xi_+}\right) \\ &\approx 1 - F_{\mathcal{W}2}\left(\frac{t_G-\mu_+}{\xi_+}\right) + \frac{1}{2NK}\left(\frac{\mu_+}{\xi_+}\right)^2 F''_{\mathcal{W}2}\left(\frac{t_G-\mu_+}{\xi_+}\right). \end{aligned}$$

The above equation can be numerically inverted to find the required threshold $t_G(\alpha)$.

3.2.3.2 *Detection Probability*

To derive an explicit approximate expression for the detection performance of the GLRT under $\mathcal{H}_1$, it is first observed that

$$\frac{1}{K}\sum_{j=1}^{K}\lambda_j = \frac{1}{K}\left[\lambda_1 + \sum_{j=2}^{K-1}\lambda_j\right]. \tag{3.61}$$

Then, the GLRT can be rewritten as

$$\lambda_1 > \tilde{t}(\alpha)\frac{\sum_{j=2}^{K}\lambda_j}{K-1} \tag{3.62}$$

with

$$\tilde{t}(\alpha) = \frac{K-1}{K - t_{GLRT}(\alpha)} t_{GLRT}(\alpha). \tag{3.63}$$

Assuming the presence of a sufficiently strong signal (asymptotically, $\rho > \frac{1}{\sqrt{KN}}$), the largest sample eigenvalue is (with high probability) due to a signal whereas the remaining eigenvalues, $\lambda_2, \ldots, \lambda_K$, are due to noise. Let

$$Z \triangleq \frac{1}{K-1}\sum_{j=2}^{K}\lambda_j \tag{3.64}$$

denote their mean. As discussed in [60] (Eq.12), asymptotically in N, the random variable Z is Gaussian distributed with variance $O\left(\frac{1}{N(K-1)}\right)$, and with a mean value that is slightly biased downwards:

$$\mathbb{E}\left[\frac{Z}{\sigma_v^2}\right] = 1 - \frac{1}{N}\frac{K\rho+1}{K\rho} + O\left(\frac{1}{N^2}\right). \tag{3.65}$$

Then recall that, by Theorem 2, λ_1/σ_v^2 is asymptotically Gaussian distributed. For a large number of sensors ($K \gg 1$), the fluctuations of Z are relatively much smaller than those of λ_1, hence (3.62) can be approximated as

$$(1+K\rho)\left(1+\frac{K-1}{NK\rho}\right) + \frac{1+K\rho}{\sqrt{N}}\eta_1 > \tilde{t}(\alpha)\cdot\mathbb{E}\left[\frac{Z}{\sigma_v^2}\right], \tag{3.66}$$

where $\eta_1 \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable. Therefore,

$$P_d(\text{GLRT}) \approx Q\left[\sqrt{N}\left(\tilde{t}(\alpha)\left(\frac{1}{K\rho+1}-\frac{1}{NK\rho}\right)-\frac{K-1}{NK\rho}-1\right)\right]. \tag{3.67}$$

Note that the above analysis assumes the presence of a *single signal*, as the GLRT (2.22) is specifically designed for this scenario.

3.3 Single Signal Detection: Impact of Noise Level Knowledge

This section is focused on the case of single-signal detection and compares the performance of different eigen-based tests under this condition. A derivation is presented based on a Neyman-Pearson test formulation, considering the cases of known and unknown noise variance (σ_v^2) and then the impact of *a priori* noise level knowledge is discussed.

3.3.1 *Neyman-Pearson Test Formulation*

3.3.1.1 *Known Noise Variance*

If $P = 1$ and σ_v^2 is known, both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are simple. When testing a simple hypothesis against a simple alternative, in general, the most powerful test is given by the Neyman-Pearson likelihood ratio [89]. In the considered scenario, with unknown channel vector $\boldsymbol{h}$, the eigenvalues of the sample covariance matrix $\boldsymbol{R}$ are sufficient statistics for the NP test (see [84]–p.11, and [61]–Sec.III-A), which can be written as

$$\text{LRT} = \frac{p(\lambda_1,\cdots,\lambda_K|\mathcal{H}_1)}{p(\lambda_1,\cdots,\lambda_K|\mathcal{H}_0)}. \tag{3.68}$$

In the asymptotical regime $N \to \infty$, with given signal strength ρ and noise variance σ_v^2, the above criterion can be shown [84, 61] to depend only on the largest eigenvalue (λ_1), i.e., it reduces to Roy's largest root test [120] T_{RLRT} (2.20).

Notice that energy detection (ED), another commonly used criterion for known noise level, is suboptimal to RLRT in the sense of the Neyman-Pearson lemma. It can be written as

$$T_{ED} \triangleq \frac{1}{KN\sigma_v^2} \sum_{k=1}^{K} \sum_{n=1}^{N} |y_k(n)|^2 = \frac{\|\boldsymbol{Y}\|_F^2}{KN\sigma_v^2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Since $\|\boldsymbol{Y}\|_F^2 = \mathrm{tr}(\boldsymbol{Y}\boldsymbol{Y}^H)$, it follows that $T_{ED} = \frac{1}{K\sigma_v^2} \sum_{i=1}^{K} \lambda_i$. Therefore, asymptotically in N, ED has reduced statistical power compared to RLRT as it tests against the noise level the sum of all eigenvalues, instead of just λ_1.

3.3.1.2 *Unknown Noise Variance*

When σ_v^2 is unknown, $\mathcal{H}_0$ and $\mathcal{H}_1$ are composite hypothesis and the Neyman-Pearson lemma does not apply. A common procedure is the GLRT, obtained from

$$\mathrm{GLRT} = \frac{\sup_{\boldsymbol{h},\sigma_v^2} p(\boldsymbol{Y}|\mathcal{H}_1)}{\sup_{\sigma_v^2} p(\boldsymbol{Y}|\mathcal{H}_0)}. \tag{3.69}$$

The GLRT procedure, that in the considered model leads to T_{GLRT} (see [10], Sec. II), can be considered optimal in a combined Neyman-Pearson/Bayesian sense [83].) Note that, since

$$\frac{1}{T_{GLRT}} = \frac{\sum_{i=1}^{K} \lambda_i}{\lambda_1} = 1 + \frac{\sum_{i=2}^{K} \lambda_i}{\lambda_1},$$

the GLRT is equivalent (up to a nonlinear monotonic transformation) to

$$T_{GLRT'} = \frac{\lambda_1}{\frac{1}{K-1} \sum_{i=2}^{K} \lambda_i}. \tag{3.70}$$

The denominator of $T_{GLRT'}$ is the maximum likelihood (ML) estimate of the noise variance assuming the presence of a signal [149], hence the GLRT can be interpreted as a largest root test with an estimated $\hat{\sigma}_v^2$ instead of the true (unknown) σ_v^2. Clearly, the GLRT is equivalent to the ERD if $K = 2$. On the contrary, the ERD is more general than the GLRT as it is applicable for any value of $P < K$.

3.3.2 *Impact of Noise Level*

3.3.2.1 *Performance Gap between RLRT and GLRT*

Next, the expressions for the detection performance of RLRT and GLRT are compared in presence of a single signal. The following result quantifies the performance gap in terms of SNR required by the two methods to achieve the same detection probability at a fixed false alarm rate.

Proposition 1. Let $\rho_R(\alpha,\beta)$ and $\rho_G(\alpha,\beta)$ be the signal-to-noise ratios required by RLRT and GLRT to achieve the same detection probability $P_d = \beta$ at the same false alarm rate α. Then, for sufficiently large K and N,

$$\frac{\rho_G(\alpha,\beta)}{\rho_R(\alpha,\beta)} \approx \frac{\tilde{t}_G(\alpha)}{t_R(\alpha)} + \frac{1}{K\rho_R}\frac{\tilde{t}_G(\alpha)-t_R(\alpha)}{t_R(\alpha)}. \tag{3.71}$$

Proof. The result follows by setting (3.39)=(3.67), and neglecting $O(1/N)$ terms. □

3.3.2.2 *Hybrid RLRT*

Given the potentially large performance gap due to not knowing the noise level, a natural question is whether this gap can somehow be overcome. In this section a novel hybrid approach is proposed, whereby the noise is estimated using *auxiliary* noise-only samples. To this end, assume that the noise variance is *constant* over adjacent time slots. This is a reasonable assumption in several CR scenarios, in case of thermal noise and constant or slowly-varying temperature. In this case, instead of estimating the noise within the sensing slot, an independent noise estimation is performed in auxiliary "noise-only" $\mathcal{H}_0$ slots where the primary signal is known to be absent[1]. In each such slot the noise variance can be estimated by using all the sample eigenvalues or, equivalently[2], by measuring the average sample energy:

$$\hat{\sigma}_v^2 = \frac{1}{K}\sum_{i=1}^{K}\lambda_i = \frac{1}{KN}\sum_{k=1}^{K}\sum_{n=1}^{N}|y_k(n)|^2. \tag{3.72}$$

Under the assumption that noise variance is (locally) constant, the estimation can be averaged over S successive noise-only slots:

$$\hat{\sigma}_v^2(S) = \frac{1}{KNS}\sum_{i=1}^{S}\sum_{k=1}^{K}\sum_{n=1}^{N}|y_k(n,i)|^2, \tag{3.73}$$

where the notation $y_k(n,i)$ stands for n-th sample in i-th time slot. The estimate $\hat{\sigma}_v^2(S)$ can then be used to replace σ_v^2 in the RLRT statistic. The resulting method is denoted *Hybrid-RLRT* (HRLRT). Its test statistic is

$$T_H \triangleq \frac{\lambda_1}{\hat{\sigma}_v^2(S)}. \tag{3.74}$$

As shown next, the performance of the HRLRT is superior to that of the GLRT and nearly achieves that of RLRT after a small number of auxiliary slots (e.g., $S = 1$ or 2).

1 In absence of prior knowledge, $\mathcal{H}_0$ slots can be identified by a GLRT with a suitably low false alarm rate.

2 By definition of Frobenius norm: $\|\boldsymbol{Y}\|_F^2 = \mathrm{tr}(\boldsymbol{Y}\boldsymbol{Y}^H)$.

Next, analytical results are provided to characterize the performance of the proposed detector. Define for the HRLRT asymptotic false-alarm probability,

$$\overline{P_{fa}}(\text{HRLRT}) \triangleq \lim_{K,N\to\infty,K/N=c} \Pr\left(T_H > t_H|\mathcal{H}_0\right) \tag{3.75}$$

and detection probability

$$\overline{P_d}(\text{HRLRT}) \triangleq \lim_{K,N\to\infty,K/N=c} \Pr\left(T_H > t_H|\mathcal{H}_1\right). \tag{3.76}$$

These quantities can be expressed as follows.

Proposition 2. Let $D \triangleq 2KNS$. in the joint limit $K, N \to \infty$ with $K/N \to c \in (0,\infty)$, the asymptotical HRLRT false alarm and detection probabilities are given by

$$\overline{P_{fa}}(\text{HRLRT}) = 1 - F_0(t_H), \qquad \overline{P_d}(\text{HRLRT}) = 1 - F_1(t_H), \tag{3.77}$$

where

$$F_0(t) \triangleq \sqrt{\frac{D}{4\pi}} \int_{-\infty}^{\infty} F_{\mathcal{W}2}\left(\frac{tx-\mu_+}{\xi_+}\right) e^{-D(x-1)^2/4} dx, \tag{3.78}$$

and where $F_1(t)$ is the distribution corresponding to the density:

$$f_1(z) \triangleq \frac{b(z)\cdot c(z)}{a^3(z)} \frac{\sqrt{D}}{2\sqrt{\pi}\sigma_1} \operatorname{erf}\left(\frac{b(z)/a(z)}{\sqrt{2}}\right) + \frac{\sqrt{D/2}}{a^2(z)\cdot\pi\sigma_1} e^{-\frac{1}{2}\left(\frac{\mu_s^2}{\sigma_1^2}+\frac{D}{2}\right)} \tag{3.79}$$

with

$$a(z) = \sqrt{\frac{1}{\sigma_1^2}z^2 + \frac{D}{2}}, \quad b(z) = \frac{\mu_s}{\sigma_1^2}z + \frac{D}{2}, \quad c(z) = e^{\frac{1}{2}\frac{b^2(z)}{a^2(z)} - \frac{1}{2}\left(\frac{\mu_s^2}{\sigma_1^2}+\frac{D}{2}\right)} \tag{3.80}$$

with μ_+ and ξ_+ given by (3.6), (3.37), μ_s given by (3.27), and $\sigma_1 \triangleq N^{-1/2}\nu_s$ (3.28).

Proof. The variable $X_D \triangleq \hat{\sigma_v}^2(S)/\sigma_v^2$ is the sum of the squares of $D = 2KNS$ real-valued Gaussian random variables, hence it has a chi-square distribution with D degrees of freedom. Next, under $\mathcal{H}_0$, λ_1/σ_v^2 is asymptotically distributed according to the second-order Tracy-Widom law with centering and scaling parameters μ_+ and $N^{-2/3}\nu_+ = \xi_+$. Therefore, under the null hypothesis $\mathcal{H}_0$,

$$\Pr[T_H < t_H] = \Pr\left[\frac{\lambda_1/\sigma_v^2}{X_D} < t\right] = \int F_{\mathcal{W}2}\left(\frac{tX_D - \mu_+}{\xi_+}\right) p(X_D = x)dx$$

where $p(X_D)$ is the density of the random variable X_D. For large D, X_D converges to a Gaussian variable: $\frac{\hat{\sigma}_v^2(S)}{\sigma_v^2} \xrightarrow{D} \mathcal{N}(1, \frac{2}{D})$, which yields $\overline{P_{fa}}$ in (3.77).

Under $\mathcal{H}_1$, λ_1/σ_v^2 is asymptotically distributed as a normal variable $\mathcal{N}(\mu_s, \sigma_1^2)$ with parameters μ_s (3.27) and $N^{-1/2}\nu_s = \sigma_1$ (3.28). The PDF of the ratio between two normal random variables can be expressed in closed-form [45] and is given by (3.79). Then, $P_d(\text{HRLRT}) = 1 - F_1(t_H)$. □

The above expression of $\overline{P_{fa}}(\text{HRLRT})$ can be inverted numerically to find the decision threshold $t_H(\alpha)$.

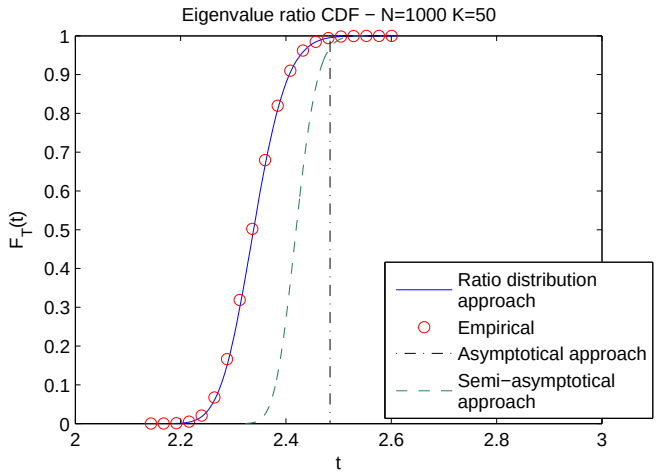


Figure 3.1: Eigenvalue ratio CDF obtained using the novel ratio-distribution approach vs. empirical CDF, asymptotical approach [17], and semi-asymptotic approach [167]. $N = 1000$, $K = 50$.

3.4 SIMULATION RESULTS

3.4.1 *ERD Performance*

Fig. 3.1 represents the eigenvalue ratio CDF resulting from the novel analytical approach (Sec. 3.2.2.1) compared to the empirical distribution, computed by Monte-Carlo simulation, and to the distributions obtained using previously proposed approaches: the asymptotical approach of [17], and the semi-asymptotic approach of [167]. The number of samples is set to $N = 1000$ and the number of cooperating receivers to $K = 50$. The analytical CDF, based on the limiting eigenvalue ratio distribution, matches with the empirical data, whereas the asymptotic one (which is simply a step function) and the semi-asymptotic one are very unbalanced because the considered parameters ($N = 1000$ samples and $K = 50$ receivers), although large, are still far from the asymptotical region. From the detector's point of view, this means that neither the asymptotic nor the semi-asymptotic approach allow to set the decision threshold correctly according to the target P_{fa}.

Fig. 3.2 provides a performance comparison of the considered eigenvalue-based detectors, plus the traditional energy detector using a cooperative equal gain combining scheme [71]. This type of plot, commonly used for signal detection and called Complementary-ROC, represents the achievable probability of missed detection $P_{md} = P(T < \gamma|\mathcal{H}_1)$ vs. the target P_{fa}. The simulation parameters are again $N = 1000$ and $K = 50$; the SNR under $\mathcal{H}_1$ is $\rho = -21$ dB. Such low values of SNR are typically used to evaluate detectors in critical conditions (e.g., in the case of "hidden node"). For energy detection, a noise uncertainty of 0.25 dB is assumed, whereas ERD is insensitive to noise power uncertainty. The ROC plot shows that the novel ratio-

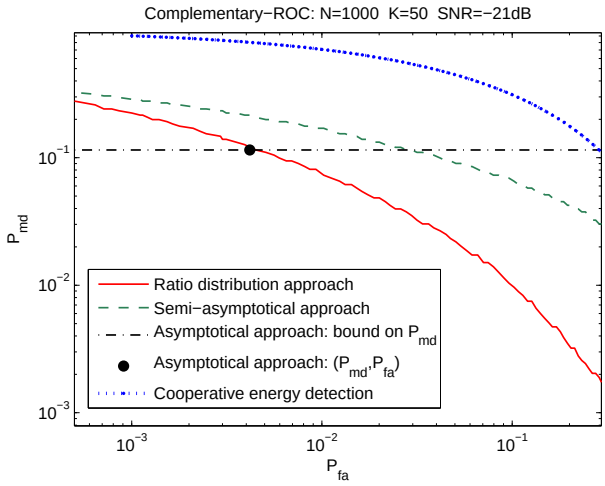


Figure 3.2: Complementary ROC: novel ratio-distribution approach vs. asymptotical approach, semi-asymptotic approach, and energy detection (with noise uncertainty). $N = 1000$, $K = 50$, SNR$= -21$dB.

distribution threshold provides lower probabilities of missed detection than the other approaches for any given probability of false alarm. Since the new algorithm uses a nearly-exact distribution, it allows to choose the lowest possible threshold for a given target P_{fa}, i.e., to obtain the minimum value of P_{md}.

For instance, given a target P_{fa} of 10^{-1}, the novel approach provides a P_{md} of $1.0 \cdot 10^{-2}$, while the semi-asymptotic approach would give $6.5 \cdot 10^{-2}$. The asymptotical approach, as previously mentioned, does not allow any control of P_{md} vs. P_{fa} since the threshold is fixed. The pair of (P_{fa}, P_{md}) it achieves is represented by a dot in the figure, at $(4 \cdot 10^{-3}, 1.15 \cdot 10^{-1})$; this value of $P_{md} = 1.15 \cdot 10^{-1}$ is a lower bound that cannot be improved regardless of the target P_{fa}, as highlighted by the straight dashed line.

Figures 3.3 and 3.4 show the convergence of the empirical CDFs to the analytical CDFs of the ERD test statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$. Four different couples of $\{N, K\}$ have been considered while keeping their ratio c fixed at 0.1. Remarkably, even though the CDFs are asymptotic, they still provide an accurate approximation of the empirical distributions even for low K and N.

In the case $\mathcal{H}_0$, as N and K increase the CDF tends to a step function, because the largest and the smallest eigenvalues converge (almost surely) to the values $\mu_+(c)$ and $\mu_-(c)$, respectively; the variance, on the contrary, depends also on N (it gets smaller for larger N).

For the case $\mathcal{H}_1$, a scenario with $P = 1$ is considered. To make the comparison more evident, t_1 is fixed instead of the SNR (ρ and t_1 are linked by a factor K, so they can not remain both constant with different K). In particular the value $t_1 = 2$ is chosen (larger than the critical value $1 + \sqrt{c} = 1.3162$ for all the considered couples of $\{N, K\}$). Similarly as in the previous case, the

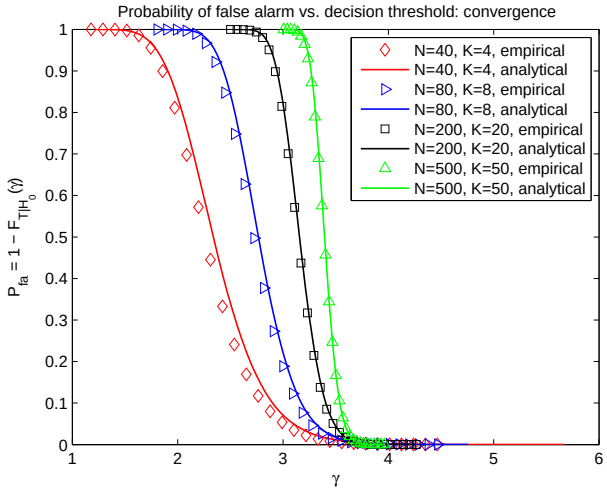


Figure 3.3: False-alarm probability: convergence, for a fixed $c = K/N = 0.1$.

CDFs turn out to converge to a step function corresponding to the almost sure asymptotical limits of the eigenvalues.

3.4.2 *RLRT and GLRT Detection Probability*

In Fig. 3.5 the accuracy of the analytical expressions of RLRT (3.39) and GLRT (3.67) detection probability are evaluated for realistic number of sensors K and samples N. The first panel refers to the case of $(K, N) = (6, 80)$ and the second panel to $(K, N) = (10, 200)$. In both cases, the false alarm rate is $\alpha = 0.5\%$. The analytical expressions of the detection probability based on RMT provide an excellent match with simulation results.

3.4.3 *SNR Gap Between RLRT and GLRT*

Fig. 3.6 illustrates the SNR gap in dB between the GLRT and RLRT. The analytical expression (3.71) is compared to simulation results and proves to be very accurate also for finite, realistic values of K and N. In accordance to the theoretical formula (3.71), the performance gap increases for (i) small number of sensors and (ii) low signal strength. This highlights the impact of noise variance estimation for eigenvalue-based sensing. As shown in the figure, for typical settings with a small number of sensors, the gap may be significant, up to 6 dB for the values considered in the figure ($K = 3$, $\rho_{RLRT} = -15$ dB). As the number of sensors increases, the noise estimate is more accurate and the performance gap between GLRT and RLRT becomes negligible. In the limit $K \to \infty$, they achieve the same detection performance, (as expected from theory, [10]). Nonetheless, even for $K = 10$ with $\rho_R = -10$ dB the gap is still of the order of 1 dB.

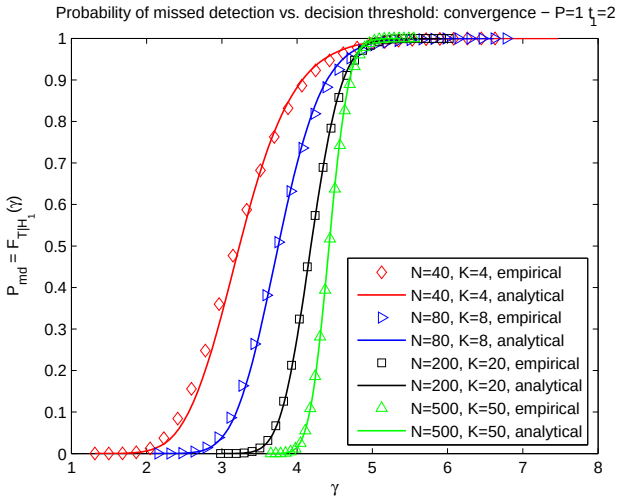


Figure 3.4: Missed-detection probability: convergence, for a fixed $c = K/N = 0.1$. $P = 1$, $t_1 = 2$.

3.4.4 *HRLRT Performance*

HRLRT simulation results are presented in Fig. 3.7. The first panel shows the accuracy of the expressions of HRLRT false-alarm and detection probability given by Proposition 3, again for realistic values of K and N. The second panel shows the performance of the HRLRT when the noise variance is independently estimated over an increasing number S of auxiliary noise-only slots. Note that the HRLRT substantially improves the performance of GLRT already for $S = 1$ or $S = 2$. At $S = 5$ the performance of HRLRT is nearly the same as that of RLRT.

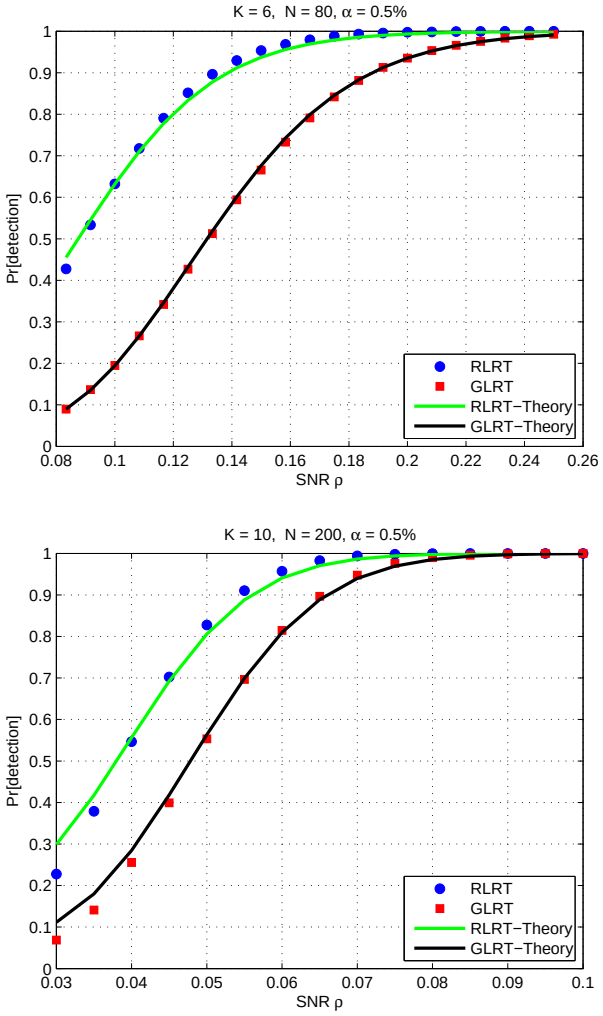


Figure 3.5: Comparison of simulated and analytical detection performance curves of RLRT and GLRT methods for different values of K, N.

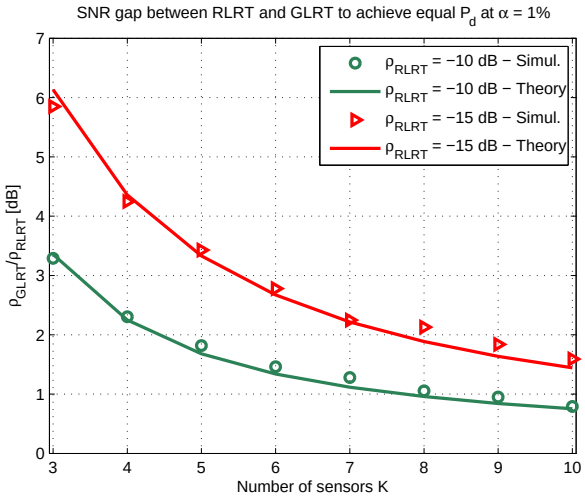


Figure 3.6: SNR gap for GLRT to achieve the same P_{det} as RLRT, as a function of K for different levels of signal strength (ρ_{RLRT}). Simulation results vs. analytical approximation (3.71). Fixed false alarm rate $\alpha = 1\%$, $N = 100$.

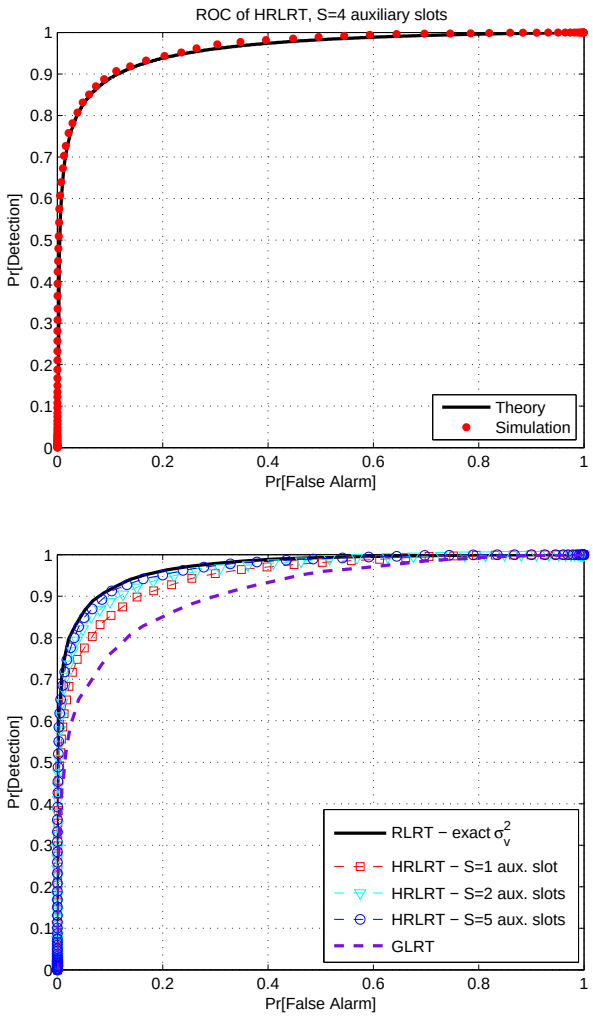


Figure 3.7: ROC curves for HRLRT. Left: validation of the expressions of false-alarm and detection probabilities. Right: comparison with GLRT and RLRT. $N = 80$, $K = 6$, $S = 4$, $\rho = -10$dB.

4

CASE STUDY: SPECTRUM SENSING IN LOW-POWER NETWORKS

4.1 INTRODUCTION

Networks of low-power short-range wireless devices, such as wireless sensor networks (WSNs) are used in a variety of applications, including environmental and industrial monitoring, surveillance, traffic control, healthcare, domotics, and military applications. In the future, it is envisaged that such devices will be communicating through the Internet protocol (IP), in particular IPv6, participating in a wide-scale network of interconnected objects, called the Internet of things (IoT) [139]. The increasing number and dense deployment of low-power wireless devices involves potential coexistence issues. Specifically, WSN devices are subject to interference for the following reasons.

- Most of current WSNs, e.g., based on IEEE 802.15.4 and Zigbee, do not implement advanced frequency selection features. Therefore, if interference occurs on the operating channel, communication can be seriously degraded.
- WSNs communicate in unlicensed bands, typically the 2.4 GHz ISM band, which is shared by other technologies, including IEEE 802.11 Wi-Fi and Bluetooth.
- WSN devices use relatively low transmit power, hence they are potentially vulnerable to several sources of interference. Wi-Fi, due to its high transmission power, represents today the main source of interference for WSNs.

For the above reasons, fixed frequency assignment does not provide reliable network operation. A possible solution is offered by adopting dynamic spectrum allocation (DSA) mechanisms, or *frequency agility (FA)*, thus extending the CR paradigm [81] to low-power networks.

In this chapter, several aspects related to the implementation of frequency agility (FA) features in WSNs are presented and analyzed. The resulting FA-WSNs are spectrum-aware, i.e., they constantly monitor the occupancy state of all possible channels, and reallocate themselves to the best available channel in real-time when interference is detected on the current operating channel.

The chapter is organized as follows. Sec. 4.2 introduces a mathematical system model. Sec 4.3 illustrates spectrum sensing and channel selection strategies, and investigates the tradeoff between sensing time and achieved

throughput by numerical results. Sec. 4.4 describes a real FA-WSN prototype and presents experimental results to evaluate its performance. Sec. 4.5 gives mathematical results about the detection of discontinuous signals (of interest for FA-WSN).

4.2 SYSTEM MODEL

The key capabilities of a FA-WSN are: (i) *spectrum sensing*, i.e., monitoring the occupancy state of the spectrum of interest, including the current operating channel as well as other potential channels; (ii) selection of a new channel whenever the current one is undergoing excessive interference; and (iii) automatic reallocation of the network to the new selected channel. Next, a network operation model and a simple mathematical model for the interfering signal are introduced.

4.2.1 *Network Operation Model*

The spectrum sensing and channel selection framework can be formalized as follows. The available spectrum consists of a set of N_c channels, denoted as $\mathcal{C}$. In IEEE 802.15.4, the ISM band contains $N_c = 16$ channels: $\mathcal{C} = \{c_{11}, \ldots, c_{26}\}$. Ideally, the operation of the proposed FA-WSN can be divided into three phases:

1. *Spectrum sensing phase*, of duration T_s, where the network (more specifically, a subset of designated spectrum sensing nodes) scans the spectrum in order to estimate the occupation state of each channel. The test statistic for spectrum sensing consists of N_s samples per channel, taken with a sampling period of t_s (depending on hardware), such that

$$T_s = N_s N_c t_s. \tag{4.1}$$

2. *Processing phase*, of duration Δ, where the spectrum sensing information is processed, a "best channel" c^* is selected and, if harmful interference is detected on the current channel, a channel switch procedure towards c^* is initiated.

3. *Communication phase*, of duration T_c, where nodes communicate on the current operational channel.

Let T_f be the system time frame duration. According to the above model, $T_f = T_s + \Delta + T_c$.

Note: The duration of the processing phase Δ can be considered negligible if no channel switch is performed, whereas it is significantly longer in case of a channel switch procedure. In the considered system, it is assumed that

normal network operation is suspended until the network has completed reallocation on the new channel c^*. Clearly, frequency reallocation introduces a delay but allows for increased throughput in the subsequent communication phase. This tradeoff will be investigated by simulations and experimental results.

4.2.2 *Interfering Signal Model*

Denote by $y_c(n)$ a generic sample of the interfering signal (e.g., Wi-Fi) received at time n on channel c. Denoting as $x_c(n)$ the interfering signal, define two alternative hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ such that

$$y_c(n) = \begin{cases} x_c(n) + v(n) & \text{under} \quad \mathcal{H}_1 \\ v(n) & \text{under} \quad \mathcal{H}_0 \end{cases} \tag{4.2}$$

where $v(n) \sim \mathcal{N}_C(0, \sigma_v^2)$ is circularly symmetric complex Gaussian (CSCG) noise. Let

$$P_x(c) \triangleq \mathbb{E}|x_c(n)|^2 \tag{4.3}$$

and

$$P_y(c) \triangleq \mathbb{E}|y_c(n)|^2, \tag{4.4}$$

which is equal to $P_x(c) + \sigma_v^2$ under $\mathcal{H}_1$ and to σ_v^2 under $\mathcal{H}_0$.

Denote as $p(c)$ the probability of occurrence of $\mathcal{H}_1$ for channel c; the probability of $\mathcal{H}_0$ is then $1 - p(c)$. Note that received samples are considered as independent from each other (no correlation between consecutive samples is taken into account), because the sampling time of WSN devices is limited by hardware constraints, and is typically larger than the burst duration of interfering (e.g., Wi-Fi) signals. The WSN transmit power is denoted as P_{WSN}. It is assumed that

$$\sigma_v^2 < P_{WSN} \leqslant P_x(c). \tag{4.5}$$

Typical values are: $\sigma_v^2 \approx -95$dBm, $P_{WSN} \approx -75$dBm; $P_x(c) \in [-75, -55]$dBm (for Wi-Fi). Real measurement results are reported in [97].

4.3 Interference Detection and Dynamic Channel Selection

In this section the key features of FA-WSN, i.e., spectrum sensing and channel selection, are analyzed in details, and the tradeoff between duration of the sensing phase and overall throughput achieved by the network is investigated.

4.3.1 *Spectrum Sensing*

Several methods of spectrum sensing were proposed in the CR literature, including matched filter detection [20], energy detection [138], cyclostationary feature detection [30], eigenvalue-based detection [166]. Considering the limitations of radio chip capabilities of typical WSN platforms, ED is the most suitable spectrum sensing method for WSNs. For example, the considered prototype – based on TelosB nodes [77], a very popular platform for WSNs – includes a spectrum sensing application (see [113]), developed in TinyOS [134], which performs ED by exploiting the built-in received signal strength indicator (RSSI)" functionality of the TelosB radio transceiver (CC2420, see [133]). At the end of the sensing time for a certain channel c (of duration T_s/N_c), the receiver has collected N_s energy samples,

$$\left[|y_c(1)|^2, \ldots, |y_c(N_s)|^2\right]. \tag{4.6}$$

4.3.2 *Channel Selection Criteria*

Based on the energy sample vector obtained on a given channel, different parameters can be used to evaluate the level of interference and identify the best channel. Next a capacity-based channel selection criterion is introduced, followed by two simplified criteria.

4.3.2.1 *Capacity-based Channel Selection*

The overall goal of the sensing phase is to select the channel that maximizes the expected throughput obtained in the subsequent communication phase. Similar to [67], the optimal channel can be defined as

$$c^*_{cap} = \arg\max_{c \in \mathcal{C}} \mathbb{E}[R_c(P_y)], \tag{4.7}$$

where R_c is the throughput of channel c. Assuming Gaussian signals and noise, the throughput can be expressed theoretically by the channel capacity given by

$$R_c(P_y) = \log_2\left(1 + \frac{P_{WSN}}{P_y(c)}\right). \tag{4.8}$$

$\mathbb{E}[R_c]$ can be expressed by averaging R_c over the distribution of $P_y(c)$. If $P_x(c)$ is constant over the sensing time, the average only involves the distribution of $\mathcal{H}_0/\mathcal{H}_1$, i.e.,

$$\mathbb{E}_{P_y}[R_c] = \Pr[\mathcal{H}_0] R_c(P_y(c)|\mathcal{H}_0) + \Pr[\mathcal{H}_1] R_c(P_y(c)|\mathcal{H}_1).$$

Recalling from (4.4) that $P_y(c)|\mathcal{H}_0 = \sigma_v^2$ and $P_y(c)|\mathcal{H}_1 = P_x(c) + \sigma_v^2$, it follows that

$$\begin{aligned}\mathbb{E}_{P_y}[R_c] =& (1-p(c)) \log_2\left(1+\frac{P_{WSN}}{\sigma_v^2}\right) \\ &+ p(c)\log_2\left(1+\frac{P_{WSN}}{P_x(c)+\sigma_v^2}\right). \end{aligned} \tag{4.9}$$

If $P_x(c)$ is variable (e.g., due to fast fading, or variable interferer transmit power in the sensing time, or bursty transmission), the average throughput is given by

$$\begin{aligned}\mathbb{E}_{P_y}[R_c] = \mathbb{E}_{P_x}\mathbb{E}_{P_y|P_x}[R_c] =& (1-p(c)) \log_2\left(1+\frac{P_{WSN}}{\sigma_v^2}\right) \\ &+ p(c)\mathbb{E}_{P_x}\log_2\left(1+\frac{P_{WSN}}{P_x(c)+\sigma_v^2}\right). \end{aligned}$$

In both cases, the average throughput cannot be computed explicitly by the sensor nodes, as parameters $p(c)$, $P_x(c)$, and the distribution of $P_x(c)$, are unknown. However, $\mathbb{E}_{P_y}[R_c]$ can be estimated as follows (see also [113]). Compute for each channel an *empirical* PDF, defined as a vector $[\pi_c(1), \ldots, \pi_c(M)]$, where M is the number of intervals used to quantize the range of energy samples $|y_c(n)|^2$ (4.6). Let $e(k)$ ($k = 1, \ldots, M$) be the quantized energy value associated to the k-th interval (e.g., the central value of such interval). Then, the average throughput $R(c)$ can be approximated as

$$\mathbb{E}_{P_y}[R_c] \approx \sum_{k=1}^{M} \log\left(1+\frac{P_{WSN}}{1+e(k)}\right)\pi_c(k). \tag{4.10}$$

This method based on the theoretical notion of capacity can be simplified by replacing R_c in (4.8) by approximate and more practical expressions of the throughput.

4.3.2.2 *Average Energy Channel Selection*

A possible channel selection criterion can be derived by approximating the average throughput as the throughput achieved under average interference conditions, i.e.,

$$\mathbb{E}_{P_y}[R_c(P_y)] \approx R_c(\mathbb{E}[P_y]) = \log_2\left(1+\frac{P_{WSN}}{\mathbb{E}[P_y(c)]}\right). \tag{4.11}$$

As a result of Jensen's inequality, the above approximation is an *upper bound* of the effective capacity[1]. Next it is shown that, for large number of samples, the average energy is a sufficient statistic for maximizing (4.11). Define

$$T_{ED}(c) \triangleq \frac{1}{N_s}\sum_{n=1}^{N_s}|y_c(n)|^2. \tag{4.12}$$

1 The relation between $\mathbb{E}_{P_y}[R_c(P_y)]$ and $R_c(\mathbb{E}[P_y])$ is similar to that between the capacities of a fading channel and an additive white gaussian noise (AWGN) channel.

In the limit $N_s \to \infty$, the sample average converges to the expected value, $T_{ED}(c) \to \mathbb{E}[|y_c(n)|^2]$, which from (4.2) equals $\sigma_v^2 + p(c)P_x(c)$, i.e., $\mathbb{E}[P_y(c)]$.

Then, since $\log(1 + P_{WSN}/x)$ is a monotone decreasing function of x, maximizing (4.11) is equivalent to minimizing $T_{ED}(c)$, i.e., selecting the channel with minimum average energy. Hence, define

$$c^*_{ED} = \arg\min_{c \in \mathcal{C}} T_{ED}(c). \tag{4.13}$$

Note that the ED criterion is simpler than (4.7) and does not require computation of the empirical PDF

4.3.2.3 *Outage Probability Channel Selection*

Another criterion can be derived by approximating the WSN throughput based on the following model: the transmission of a packet fails whenever the interference is above a predefined level (denoted by γ), and it is successful otherwise. For example, the CC2420 transceiver defines the level of $\gamma = -77$dBm as the threshold for "clear channel assessment", which is performed before every transmission. Motivated by the above consideration, $R_c(P_y)$ can be approximated by the following step function:

$$R_c(P_y) \approx \begin{cases} \log_2\left(1 + \frac{P_{WSN}}{\sigma_v^2}\right) & \text{if} \quad P_y < \gamma \\ 0 & \text{if} \quad P_y > \gamma \end{cases} \tag{4.14}$$

That is, P_y is approximated as $P_y \approx \sigma_v^2$ whenever P_y is lower than γ, and $P_y \approx \infty$ when larger than γ. The first event corresponds to absence of interfering signal ($\mathcal{H}_0$) or presence of signal with negligible power (low values of P_x); the second case reflects the fact that any signal such that $P_y > \gamma$ causes serious interference issues, hence the achievable WSN throughput is close to zero.

Based on the above approximation, one can write

$$\begin{aligned} \mathbb{E}_{P_y}[R_c] &= \int_0^\infty R_c(P_y) f_{P_y(c)}(P_y) dP_y \\ &\approx \log_2\left(1 + \frac{P_{WSN}}{\sigma_v^2}\right) \int_0^\gamma f_{P_y(c)}(P_y) dP_y \\ &= \log_2\left(1 + \frac{P_{WSN}}{\sigma_v^2}\right) \Pr[P_y(c) < \gamma]. \end{aligned}$$

Since $\log_2\left(1 + \frac{P_{WSN}}{\sigma_v^2}\right)$ is a constant, maximizing the above expression amounts to selecting the channel with minimum

$$P_{out}(c) \triangleq \Pr[P_y(c) > \gamma], \tag{4.15}$$

which is defined *outage probability*.

The outage probability can be estimated from the empirical PDF for any threshold γ corresponding to one of the quantization levels. If γ is fixed,

the outage probability can be computed simply by counting the number of samples that are larger than γ. Thus, $P_{out}(c)$ is approximated as $P_{out}(c) \approx T_{out}(c)$, with

$$T_{out}(c) \triangleq \frac{1}{N_s} \sum_{n=1}^{N_s} \mathbb{I}\{|y_c(n)|^2 > \gamma\}, \tag{4.16}$$

where $\mathbb{I}\{x\} = 1$ if x is true. The best channel according to the outage probability is then selected as

$$c^*_{out} = \arg\min_{c \in \mathcal{C}} T_{out}(c). \tag{4.17}$$

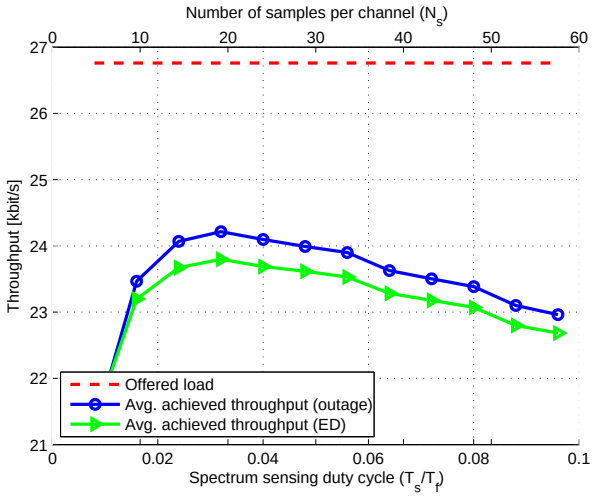
Note that this channel selection criterion is particularly appropriate in cases where P_x is not constant (e.g., fast fading or bursty signal), hence P_y exhibits significant fluctuations.

4.3.3 *Sensing-Throughput Tradeoff*

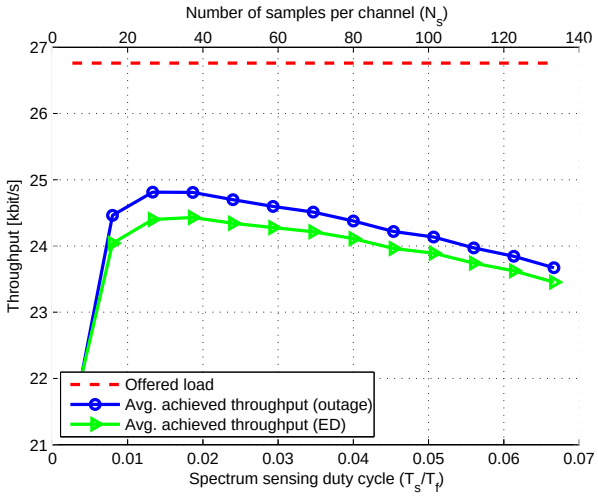
The following simulation results are aimed at evaluating numerically the tradeoff between spectrum sensing duration and throughput, and at investigating how channel selection criteria (average energy or outage probability) affect performance. A WiFi-like interference model is adopted, as discussed in Sec. 4.2. Assume $N_c = 16$ potentially available channels, with i.i.d. occupation probabilities $p(c)$ uniformly distributed in $(0, 1)$ for each channel and each frame. The interfering signal power $P_x(c)$ varies on a sample-by-sample basis, according to a log-normal distribution with mean -65dBm and standard deviation of 5dB to take into account multipath propagation. The noise level is -90dBm. Spectrum sensing is performed by a WSN node with sampling period $t_s = 0.5$ms and processing time is assumed as negligible ($\Delta = 0$). In the communication phase, the throughput is calculated assuming that a packet is successfully transmitted if the interference power on the channel is lower than $\gamma = -75$dBm. WSN packets are of size $l_p = 98$ bytes and are transmitted at a rate of $t_p = 29.3$ms, resulting in an offered load of 26.7kb/s.

Given this setting, the achieved throughput as a function of the sensing time is shown in Fig. 4.1 for two different values of frame duration. The sensing time is expressed both in terms of absolute number of samples (upper axis) and "duty-cycle", i.e., relative duration of the sensing phase over the frame time (lower axis). Results indicate that: *(i)* an optimal sensing-throughput value is clearly visible, in agreement with the results of [67]; *(ii)* the optimal number of samples is relatively low: ~ 20 for $T_f = 5$s, ~ 30 for $T_f = 15$s; *(iii)* as the frame duration increases, the achieved throughput gets closer to the offered load; *(iv)* outage probability selection slightly outperforms the average energy criterion.

The exact location of the optimum in terms of number of samples depends on several system parameters and assumptions, including above all



(a) $T_f = 5s$.



(b) $T_f = 15s$.

Figure 4.1: Simulated throughput vs. sensing time. Wi-Fi traffic, i.i.d. samples with random $p(c)$ and $P_x(c)$ for different channels.

the ideal model of i.i.d. samples. More interestingly, the introduction of spectrum sensing provides a remarkable throughput gain: for example, in Fig. 4.1(b), the achievable throughput with spectrum sensing enabled nearly reaches 25kb/s, whereas the throughput obtained without spectrum sensing is about 21kb/s. These preliminary results suggest that the adoption of FA features in WSNs can potentially lead to a significant performance improvement, as it will be confirmed in the next section, where the performance of a real FA-WSN is analyzed.

4.4 REALIZATION OF FREQUENCY-AGILE WSN

This section concentrates on implementation of FA features in WSNs based on IPv6.The proposed model encompasses (i) the network architecture enabling FA capabilities, (ii) FA extensions to a generic IP stack, and (iii) FA-related system operations and their integration with regular WSN applications. All the design aspects take into consideration the limited resources of WSN nodes or similar devices.

4.4.1 *FA Network Architecture*

Fig. 4.2 shows the reference FA architecture. The network is divided into *clusters* to improve scalability. Inside a cluster, nodes can be assigned different *roles*:

- The *Cluster Head (CH)* connects the cluster to the *Edge router* (E), and hence to the network backbone.
- *Spectrum Sensing Nodes (SSNs)* are capable of periodically performing energy detection on each of the available channels. SSNs collect energy samples as in (4.6), where such data can be used to determine the spectrum occupancy state. SSNs can completely dedicate their resources to spectrum sensing, or also contribute to normal network operations after sensing (*Data-SSNs*).
- *Non-SSNs* are ordinary network nodes that carry out application-related tasks, and do not participate in spectrum sensing.

Among the deployed nodes in an FA network, the number of SSNs is highly dependent on the application specifications, e.g., the number of available channels, the required detection responsiveness, and the area covered by the network. For instance, in applications where it is essential to maintain a high throughput, the number of deployed SSNs may be equal to the number of available channels, with each SSN set with a high sensing duration, thus ensuring maximum responsiveness and reliable detection. On the contrary, in less demanding applications, a couple of SSNs would suffice to periodically provide spectrum state information.

Note that the proposed network architecture only applies to FA features, and does not necessarily have to coincide with the WSN architecture in normal operation mode.

Spectrum sensing information is communicated to an entity called the *Frequency Agility Manager (FAM)*. The FAM processes the spectrum sensing data collected from the network and determines the spectrum occupancy state. If the FAM detects that a critical level of interference is present in the operating channel (OC), it generates an OC-switch command to reallocate the network to the best available channel. In the current implementation, the

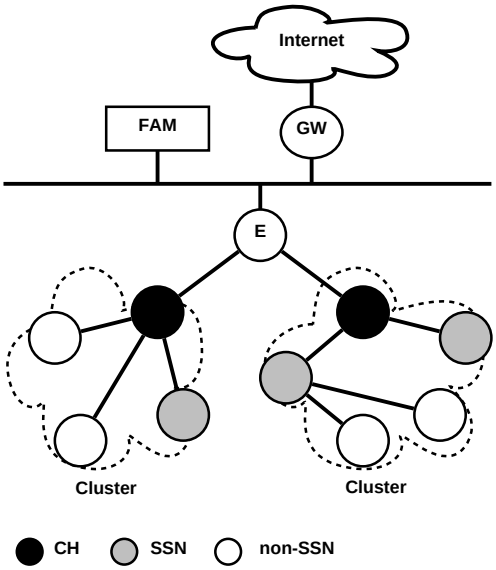


Figure 4.2: Reference network architecture to enable FA features.

FAM is a centralized entity (physically located in the edge router); distributed implementation over a number of nodes is possible as well.

4.4.2 *Performance Evaluation of FA-IPv6-WSN*

Now the performance of dynamic channel selection in the above described FA-IPv6-WSN prototype is analyzed. The goal is to evaluate the throughput achieved with and without FA features, i.e., spectrum sensing and frequency reallocation, in presence of interference. The experimental test-bed is composed of a cluster containing three SSNs: two SSNs completely dedicated to energy detection, and one Data-SSN. All SSNs are connected to a CH which controls the data between its cluster and an edge router that is connected to a Control Center (cf. Fig. 4.2). During each time frame, the Data-SSN performs spectrum sensing on one channel (indicated by the CH for an interval T_s and then sends data (UDP packets) to the edge router for the rest of the frame time. In the meanwhile, each of the two SSNs performs spectrum sensing on other channels (again, indicated by the CH) for the same interval T_s, and provides an SSN-Report every one second. Note that the CH manages the spectrum sensing operation so as to avoid duplicate SSN-Reports for a channel. The Control Center then counts the number of data packets received from the Data-SSN and calculates the average achieved throughput.

The introduction of the two SSNs beside the Data-SSN has the purpose to ensure a reasonably fast detection of the whole spectrum while keeping a relatively low spectrum sensing duty-cycle. Thus, the experimental setup differs from the ideal model presented in Sec. 4.2.1 in that spectrum sensing is performed on one channel per frame, instead of all channels in the sensing

time. The rationale for such implementation is that every channel change for the sensing node involves a delay that may be comparable to T_s, especially when the number of samples per channel is low. For this reason, scanning 16 channels within each frame is not convenient in practice.

Then, in order to generate random, time-varying interference conditions on the entire spectrum with no correlation between different channels, *jammer nodes* are used. Jammers are IEEE 802.15.4 devices, programmed so as to transmit interference at a rate of 72kb/s on a given channel for a certain amount of time, and then move to another channel. Thus, multiple operating jammers allow to emulate the desired interference conditions, which would not be feasible using Wi-Fi.

Based on this setting, first the sensing throughput-tradeoff is evaluated, similar to Sec. 4.3.3, and then the performance of a FA-enabled WSN is compared against that of a regular WSN in terms of throughput and packet loss rate under different interference conditions. The throughput is measured as the number of packets successfully transmitted from the Data-SSN to the edge router, given an offered load of 26.7kb/s (maximum achievable rate without interference).

4.4.2.1 *Sensing-throughput Tradeoff*

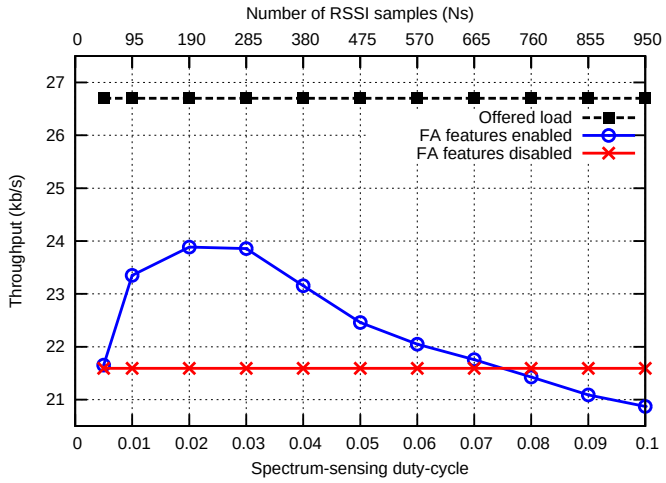


Figure 4.3: Throughput vs. sensing time: experimental results. Interference generated by 7 jammers operating on random channels.

Fig. 4.3 shows the achieved throughput as a function of the sensing duration (number of samples or duty cycle) according to the previously described implementation. Similar to Sec. 4.3.3, the sampling period is set to $t_s = 0.5$ms, the packet size to $l_p = 98$ bytes, and the packet transmission time is $t_p = 29.3$ms, and the Data-SSN frame duration is $T_f = 5$s. In contrast, one sensing window (of duration T_s) corresponds now to one channel instead of all 16 channels. Interference is generated by 7 jammers, transmitting at 0dBm and deployed in an indoor environment at different distances (5

to 10m) from the spectrum sensing node. In this experiment, jammer channels are randomly selected every 60s independently for each jammer. Results show that, when FA features are enabled, the achieved throughput increases from 21.5kb/s to 24kb/s, similar to the simulation results presented in Fig. 4.1(a). A similar tradeoff between sensing time and throughput can be observed as well: the throughput increases with the number of samples up to an optimal value, and then decreases almost linearly. In the adopted experimental configuration, the optimal number of samples per channel turns out to be around 190. A significant throughput improvement is observed not only around the optimum, but for a large range of values of N_s. Clearly, such improvement vanishes if the sensing period is too long (> 700 samples).

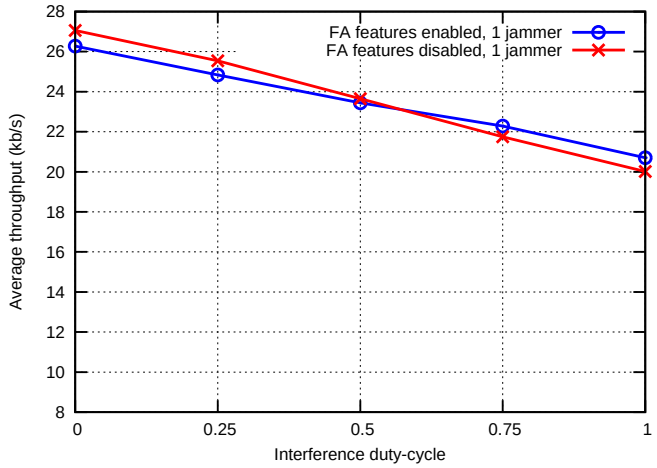
Performance has been evaluated also in terms of packet loss. The FA-IPv6-WSN provides 3% packet losses, while the non-FA WSN has a 17% loss rate.

4.4.2.2 *Comparison Between FA-IPv6-WSN and Ordinary IPv6-WSN*

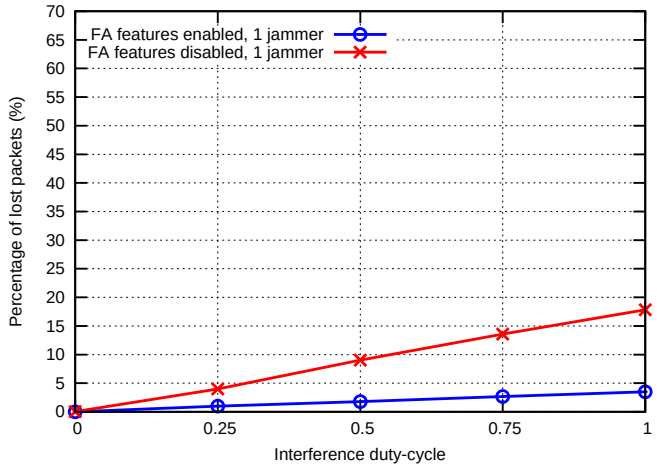
The proposed FA-IPv6-WSN and an ordinary IPv6-WSN with no FA capabilities are now compared under different interference conditions in the considered test-bed configuration. For this purpose, define the *interference duty-cycle* as the percentage of the experiment time during which the interference is present. Also, in this experiment jammers, instead of being random, are set to track and attack the network, i.e., they are equipped with FAM-scan capability such that they find the FA-WSN operating channel and generate interference on that channel. The performances of both networks are evaluated in terms of throughput and packet losses, for different interference duty-cycles as well as two interference levels: one jammer (Fig. 4.4) and two jammers (Fig. 4.5). For the FA-IPv6-WSN, the Data-SSN was set at the optimum N_s.

Fig. 4.4(a) shows the average throughput comparison. It can be seen that the network with no FA-features achieves higher throughput at lower interference duty-cycles, but its throughput drops slightly faster than the FA-WSN network as interference increases. Observe that the FA-WSN starts to outperform the ordinary network at an interference duty-cycle of 65%. Fig. 4.4(b) shows the comparison in terms of packet loss. Based on this metric, the performance improvement of the FA networks is more evident, and can be observed for any interference condition.

If the sources of interference are two jammers, as shown in Fig. 4.5(a) and Fig. 4.5(b), the performance gap between FA and non-FA WSN becomes larger. For instance, the FA-WSN is superior in terms of throughput already at an interference duty cycle of 12%. It is to be noted that UDP transmissions do not implement Acks, and therefore the improvements achieved by FA-capable networks can be even greater in systems adopting for example TCP transmissions.



(a)

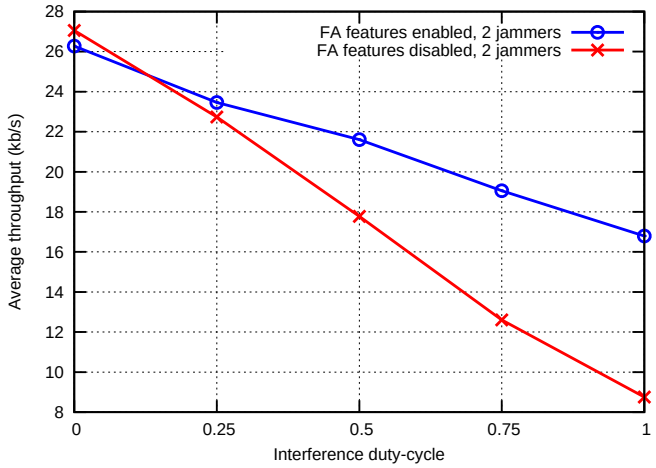


(b)

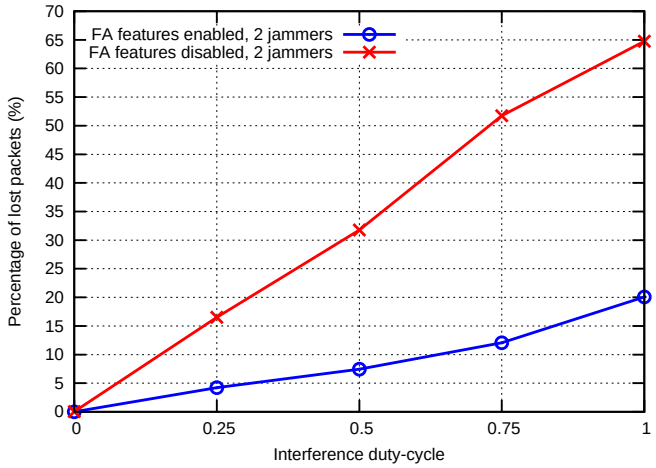
Figure 4.4: Comparison between FA-IPv6-WSN and ordinary IPv6-WSN, operating under different interference duty-cycles generated by 1 Jammer: in terms of (a)average throughput, and (b)packet losses.

4.5 DETECTION OF DISCONTINUOUS SIGNALS

This section focuses on the problem of detecting signals characterized by discontinuous occupation in the sensing window. This is the case in FA-WSN and in similar applications where, on the one hand, the sampling frequency may be limited by hardware capabilities and, on the other hand, the signal to be detected may be bursty with a burst duration comparable to or lower than the minimum sampling period. In this scenario, the performance of two detectors – the classic single-user ED and the multi-sensor, eigen-based RLRT is evaluated taking into account the signal *occupancy probability*.



(a)



(b)

Figure 4.5: Comparison between FA-IPv6-WSN and ordinary IPv6-WSN, operating under different interference duty-cycles generated by 2 Jammers: in terms of (a)average throughput, and (b)packet losses.

For completeness, the signal model is redefined below, adopting a convenient mathematical notation. Let $y(n)$ be the received base-band signal sample at time n and let N be the total number of samples collected by the receiver during the sensing time. The set of samples received during the sensing window is indicated by the $N \times 1$ vector

$$\boldsymbol{y} \triangleq [y(1) \dots y(N)] \tag{4.18}$$

In the considered scenario, two possible hypotheses exist:

- $\mathcal{H}_{\text{free}}$: signal is absent for the entire duration of the sensing window

- $\mathcal{H}_{\text{busy}}$: signal is present, with *occupancy probability* $\overline{p}_1$; the complementary absence probability is defined as $\overline{p}_0 \triangleq 1 - \overline{p}_1$

The generic received signal sample in the first case can be written as

$$y(n)|_{\mathcal{H}_{\text{free}}} = \nu(n) \tag{4.19}$$

where $\nu(n)$ is modeled as CSCG noise with zero mean and variance σ_ν^2. In the alternative case, the signal sample is

$$y(n)|_{\mathcal{H}_{\text{busy}}} = \begin{cases} \text{case 0:} & \nu(n) & \text{with probability} & \overline{p}_0 \\ \text{case 1:} & h\, s(n) + \nu(n) = x(n) + \nu(n) & \text{with probability} & \overline{p}_1 \end{cases} \tag{4.20}$$

where $s(n)$ is the transmitted signal sample, by assumption zero-mean and with variance σ_s^2, and h represents the channel (modeled as constant during the sensing window). The SNR at the single receiver is

$$\rho = \frac{|h|^2 \sigma_s^2}{\sigma_\nu^2}. \tag{4.21}$$

Within the N-sample observation interval, let N_0 and N_1 the number of samples where case 0 and case 1 occur, respectively, and define normalized occupation/absence rates

$$p_1 \triangleq \frac{N_1}{N} \tag{4.22}$$

$$p_0 \triangleq 1 - p_1 = \frac{N_0}{N} \tag{4.23}$$

Clearly, as $N \to \infty$, $p_0 \to \overline{p}_0$ and $p_1 \to \overline{p}_1$.

The occupancy rate $\overline{p}_1$ is a specific parameter of the signal to be detected. For instance, consider a Wi-Fi signal with an over-the-air packet duration of 2 ms and a packet rate of 90 pkt/s: in this case $\overline{p}_1$ is 0.18. Note that the discontinuous signal model reduces to the usual signal model adopted in most of the CR literature when $\overline{p}_1 = 1$. For simplicity, this section is focused on the case of single signal.

4.5.1 *Energy Detection*

Consider the conventional single-user ED detection scheme. The test statistic is the average energy of the received signal samples, i.e.,

$$T_{\text{ED}} \triangleq \frac{1}{N}\|\boldsymbol{y}\|^2 = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^2 \tag{4.24}$$

To analyze ED performance, it is necessary to express the PDF of the test statistic in case of discontinuous signal. The result is given by the following theorem.

Theorem 4. ***Asymptotical distribution of*** T_{ED}. *For* $N_0, N_1 \to \infty$, *the test statistic* T_{ED} *is normally distributed*

$$T_{ED} \sim \mathcal{N}\left(\mu, \zeta^2\right) \tag{4.25}$$

with

$$\mu = \sigma_\nu^2(1 + \overline{p}_1\rho) \tag{4.26}$$

$$\sigma^2 = \frac{1}{N}\sigma_\nu^4\left[1 + \overline{p}_1 g(\rho)\right] \tag{4.27}$$

where $g(\rho)$ *is a function of the SNR, specific of the signal type, e.g.,*

$$g(\rho) = \begin{cases} 2\rho & \text{if } x \text{ is complex PSK modulated} \\ \rho^2 + 2\rho & \text{if } x \text{ is complex Gaussian} \end{cases} \tag{4.28}$$

Proof. By definition of p_0 and p_1 the test statistic can be split into two separate summations:

$$T_{ED} = p_0\frac{1}{N_0}\sum_{n=1}^{N_0}|\nu(n)|^2 + p_1\frac{1}{N_1}\sum_{n=1}^{N_1}|x(n)+\nu(n)|^2 \tag{4.29}$$

The first component

$$T_0 \triangleq \frac{1}{N_0}\sum_{n=1}^{N_0}|\nu(n)|^2 \tag{4.30}$$

is a sum of Gaussian random variables, so it follows a chi-square distribution with $2N_0$ degrees of freedom up to a normalisation factor:

$$\frac{2N_0}{\sigma_0^2}T_0 \sim \chi^2_{2N_0} \tag{4.31}$$

As N_0 tends to infinity, the chi-square distribution converges to a normal distribution:

$$N_0 \to \infty : T_0(y) \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right) \tag{4.32}$$

$$\mu_0 = \sigma_\nu^2, \quad \sigma_0^2 = \frac{1}{N_0}\sigma_\nu^4 \tag{4.33}$$

The other component

$$T_1 \triangleq \frac{1}{N_1}\sum_{n=1}^{N_1}|x(n)+\nu(n)|^2 \tag{4.34}$$

does not admit a closed-form PDF; however, the central limit theorem (CLT) guarantees that for $N_1 \to \infty$ it converges to a normal distribution, with mean and variance given by (cf. [67])

$$N_1 \to \infty : T_1(y) \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right) \tag{4.35}$$

$$\mu_1 = \sigma_\nu^2 + \sigma_x^2 = \sigma_\nu^2(1+\rho) \tag{4.36}$$

$$\sigma_1^2 = \frac{1}{N_1}\left(\mathbb{E}\,|x|^4 - \sigma_x^4 + \sigma_\nu^4 + 2\sigma_x^2\sigma_\nu^2\right) \tag{4.37}$$

The value of $\mathbb{E}\,|x|^4$ depends on the signal type, e.g,

- $\mathbb{E}\,|x|^4 = 2\sigma_x^4$ for complex Gaussian signals
- $\mathbb{E}\,|x|^4 = \sigma_x^4$ for complex PSK modulated signals

More cases are listed in [67]. In general, it can be easily checked that

$$\sigma_1^2 = \frac{1}{N_1}\sigma_v^4 f(\rho) \tag{4.38}$$

where $f(\rho)$ equals $2\rho + 1$ for PSK signals or $(\rho+1)^2$ for Gaussian signals. By exploiting the linearity of the sum of Gaussian random variables, T_{ED} is still (asymptotically) Gaussian with

$$\mu = \overline{p}_0\mu_0 + \overline{p}_1\mu_1 \tag{4.39}$$
$$\sigma^2 = \overline{p}_0^2\sigma_0^2 + \overline{p}_1^2\sigma_1^2 \tag{4.40}$$

which, after some algebra and letting $g(\rho) = f(\rho) - 1$, yield (4.26) and (4.27). The occupancy rates p_0, p_1 have been replaced by their respective probability values since $N_0, N_1 \to \infty$. □

By using Theorem 4, the **probability of detection** for ED under discontinuous signal assumption can be expressed as:

$$\begin{aligned} P_d &= \Pr(T_{ED} > \gamma | \mathcal{H}_{busy}) = Q\left(\frac{\gamma - \mu}{\sigma}\right) \\ &= Q\left[\sqrt{N}\left(\frac{\gamma/\sigma_v^2 - (1+\overline{p}_1\rho)}{\sqrt{1+\overline{p}_1 g(\rho)}}\right)\right] \end{aligned} \tag{4.41}$$

Notice that the formula of the false-alarm probability does not depend on p_1 and is given as usual [138, 67] by

$$P_{fa} = \Pr(T_{ED} > \gamma | \mathcal{H}_{free}) = Q\left[\sqrt{N}\left(\frac{\gamma}{\sigma_v^2} - 1\right)\right] \tag{4.42}$$

4.5.2 *Eigenvalue-based Detection*

Consider now multi-sensor detection, with K be the number of available sensors, and let h_k be the channel coefficient associated to the k-th sensor. Channels are assumed to be constant during the sensing time, but different from one receiver to another. Like in Sec. 2, a $K \times N$ received signal matrix is available:

$$\boldsymbol{Y} \triangleq [\boldsymbol{y}_1 \dots \boldsymbol{y}_K]^T \tag{4.43}$$

The sample covariance matrix is then defined as

$$\boldsymbol{R}(N) \triangleq \frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^H \tag{4.44}$$

It can be equivalently written in terms of the columns of $\boldsymbol{Y}$, denoted by $\widetilde{\boldsymbol{y}}(n)$ for n from 1 through N, as

$$\boldsymbol{R}(N) = \frac{1}{N}\sum_{n=1}^{N} \widetilde{\boldsymbol{y}}(n)\widetilde{\boldsymbol{y}}^{H}(n) \tag{4.45}$$

which, for large N, approximates the statistical covariance matrix

$$\boldsymbol{R} \triangleq \mathbb{E}\, \widetilde{\boldsymbol{y}}\widetilde{\boldsymbol{y}}^{H} \tag{4.46}$$

Let $\ell_1 \geqslant \ldots \geqslant \ell_K$ be the eigenvalues of $\boldsymbol{R}$ (without loss of generality, sorted in decreasing order), and let $\lambda_1 \geqslant \ldots \geqslant \lambda_K$ be the eigenvalues of $\boldsymbol{R}'(N)$. The RLRT statistic is $T_{RLRT} = \lambda_1/\sigma_v^2$, as defined in (2.20) It is now convenient to define a "global" SNR

$$\rho_K \triangleq \frac{\|\boldsymbol{h}\|^2\sigma_s^2}{\sigma_v^2} \tag{4.47}$$

where $\boldsymbol{h} \triangleq [h_1 \cdots h_K]^T$. The following theorem extends an important results from random matrix theory to the scenario of interest.

Theorem 5. ***Largest eigenvalue of the covariance matrix with discontinuous signals***. *Assuming that* $s(n)$ *is Gaussian and that all* K *receivers observe the same signal with occupation rate* $p_1 = N_1/N \to \overline{p}_1$*, the following results hold.*

(i) The largest eigenvalue of the statistical covariance matrix is

$$\ell_1 = \sigma_v^2(\overline{p}_1\rho_K + 1) \tag{4.48}$$

(ii) in the limit $K, N \to \infty$*, the largest eigenvalue of the sample covariance matrix converges to*

$$\lambda_1 \xrightarrow{a.s.} \sigma_v^2(\overline{p}_1\rho_K + 1)\left(1 + \frac{K-1}{N\overline{p}_1\rho_K}\right) \tag{4.49}$$

provided that

$$\rho_K > \frac{\sqrt{K/N}}{\overline{p}_1} \tag{4.50}$$

Proof. Since the occupation probability is homogeneous for all receivers, the received signal vector at time instant n can be written, respectively, as

$$\widetilde{\boldsymbol{y}}(n)|_{\mathcal{H}_{free}} = \widetilde{\boldsymbol{v}}(n) \tag{4.51}$$

or

$$\widetilde{\boldsymbol{y}}(n)|_{\mathcal{H}_{busy}} = \begin{cases} \widetilde{\boldsymbol{v}}(n) & \text{with probability } \overline{p}_0 \\ \widetilde{\boldsymbol{x}}(n) + \widetilde{\boldsymbol{v}}(n) & \text{with probability } \overline{p}_1 \end{cases} \tag{4.52}$$

with $\widetilde{\boldsymbol{x}}(n) \triangleq \boldsymbol{h}s(n)$. By linearity of the expectation operator, the covariance matrix (4.46) is

$$\boldsymbol{R} = \begin{cases} \sigma_v^2 \boldsymbol{I}_K & (\mathcal{H}_{\text{free}}) \\ \overline{p}_1 \boldsymbol{R}_{\boldsymbol{x}} + \sigma_v^2 \boldsymbol{I}_K & (\mathcal{H}_{\text{busy}}) \end{cases} \tag{4.53}$$

with $\boldsymbol{R}_{\boldsymbol{x}} \triangleq \mathbb{E}\, \widetilde{\boldsymbol{x}}\widetilde{\boldsymbol{x}}^H = \sigma_s^2 \boldsymbol{h}\boldsymbol{h}^H$. This is a rank-1 matrix, whose only non-zero eigenvalue is $\zeta = \|\boldsymbol{h}\|^2 \sigma_s^2$. Therefore, the largest eigenvalue of $\boldsymbol{R}$ is

$$\lambda_1 = \overline{p}_1 \lambda_x + \sigma_v^2 \tag{4.54}$$

that is (4.48).

The second claim of the theorem is an extension of Theorem 2 presented in Sec. 4.5.2. As discussed, e.g., in [61, 110], a signal is identifiable only if its corresponding eigenvalue is greater than $\sigma_v^2\sqrt{c}$, with $c \triangleq K/N$. In the case considered here, the signal eigenvalue turns out be $\overline{p}_1\lambda_x$ (4.54); therefore, the identifiability threshold is given by

$$\overline{p}_1 \lambda_x > \sigma_v^2 \sqrt{c} \tag{4.55}$$

as if there were a signal with "virtual" SNR $\overline{p}_1\rho$. When this condition is verified, random matrix theory results [6] ensure that, as $N, K \to \infty$,

$$\frac{\lambda_1}{\sigma_v^2} \xrightarrow{\text{a.s.}} (\rho' + 1)\left(1 + \frac{K-1}{N\rho'}\right), \tag{4.56}$$

which (replacing the SNR ρ' by $\overline{p}_1\rho$) yields (4.49). On the contrary, if the condition is not verified, the largest sample eigenvalue converges to $\sigma_v^2(1+\sqrt{c})^2$ thus making signals indistinguishable from noise. □

The above theorem provides the asymptotical mean value of the test statistic λ_1/σ_v^2. However, to characterise the detection probability, it is necessary to understand how λ_1 is distributed around the value (4.49). It is known that the fluctuations of the largest eigenvalue in spiked population models are asymptotically Gaussian; in the general case of covariance matrices constructed from homogeneous sets of data, their variance is known as well [5, 61]. Unfortunately in the case of discontinuous signals this result cannot be applied since the signal entries do not have the same variance: the result becomes more and more inaccurate as $\overline{p}_1$ departs from 1.

To derive a better approximation of the variance, the sample covariance matrix (4.44) is expressed as

$$\boldsymbol{R}'(N)|_{\mathcal{H}_{\text{busy}}} = p_0 \boldsymbol{R}_0(N_0) + p_1 \boldsymbol{R}_1(N_1) \tag{4.57}$$

where

$$\boldsymbol{R}_0(N_0) \triangleq \frac{1}{N_0} \sum_{n=1}^{N_0} \widetilde{\boldsymbol{v}}(n)\widetilde{\boldsymbol{v}}(n)^H \tag{4.58}$$

$$\boldsymbol{R}_1(N_1) \triangleq \frac{1}{N_1} \sum_{n=1}^{N_1} [\widetilde{\boldsymbol{x}}(n) + \widetilde{\boldsymbol{v}}(n)][\widetilde{\boldsymbol{x}}(n) + \widetilde{\boldsymbol{v}}(n)]^H \tag{4.59}$$

are the partial covariance matrices constructed, respectively, from the only-noise and from the signal-plus-noise samples. $\boldsymbol{R}_0$ is a Wishart matrix and $\boldsymbol{R}_1$ a standard spiked-population covariance matrix. For both these matrices, the asymptotical distribution of the largest eigenvalue can be expressed exactly: a Tracy-Widom distribution of second order [3] for the largest eigenvalue of $\boldsymbol{R}_0$, a Gaussian distribution for that of $\boldsymbol{R}_1$.

Since the variance of the only-noise component is negligible compared to the signal component, the variance of λ_1 can be approximated as the variance of the largest eigenvalue of $\boldsymbol{R}_1(N_1)$ normalised by $\overline{p}_1^2$:

$$\mathrm{Var}[\lambda_1] \approx \overline{p}_1^2 \frac{1}{N_1} \sigma_v^4 (\rho_K + 1)^2 \left(1 - \frac{K-1}{N_1 \rho_K^2}\right) \tag{4.60}$$

This approximation turns out to be very accurate for all values of ρ_K above the identifiability threshold, and for all $\overline{p}_1$ in $(0, 1)$.

Then, the **detection probability** of Roy's largest root test, under discontinuous signal assumption, is

$$P_d = \Pr(T_{RLRT} > \gamma | \mathcal{H}_{busy}) \approx Q\left(\frac{\gamma - \mu_d}{\xi_d}\right) \tag{4.61}$$

with

$$\mu_d = (\overline{p}_1 \rho_K + 1)\left(1 + \frac{K-1}{N \overline{p}_1 \rho_K}\right) \tag{4.62}$$

$$\xi_d = \frac{\overline{p}_1}{\sqrt{N_1}} (\rho_K + 1) \sqrt{1 - \frac{K-1}{N_1 \rho_K^2}} \tag{4.63}$$

Similarly as for ED, also in this case the false alarm probability is not affected by the discontinuous occupancy. Its asymptotical expression, for $K, N \to \infty$ with $c = K/N \to \overline{c}$ finite is still given by

$$P_{fa} = \Pr(T_{RLRT} > \gamma | \mathcal{H}_{free}) = 1 - F_{TW2}\left(\frac{\gamma - \mu_+}{\nu_+}\right), \tag{4.64}$$

as discussed in Sec. 3. Even though all the expressions derived in this section and in the previous one are based on asymptotic assumptions, they are accurate enough for practical purposes already for relatively small values of K and N.

4.5.3 *Numerical Results*

In the subsequent Fig. 4.6, the detection probability of the two considered methods is analyzed as a function of the SNR with a fixed $P_{fa} = 0.1\%$. Fig. 4.7 shows ROC curves, i.e., plots of P_d as a function of P_{fa}. Finally, Fig. 4.8 shows the detection probability as a function of the number of samples, for fixed P_{fa} and SNR.

The results show a good match between simulation results and the analytical model developed in this work. Even though all formulas are obtained under the assumption of $N, K \to \infty$, they turn out to be accurate for the considered values of K and N which are realistic and far from the asymptotic limits. Also, the approximate value ξ_d used to model the variance of the RLRT test statistic is consistent with the empirical distributions.

The detection probabilities computed using a "traditional" approach (i.e., assuming $p_1 = 1$) are significantly distant from the ones obtained in a discontinuous signal scenario, both for $p_1 = 0.3$ and for $p_1 = 0.7$. From the detection point of view, it means that neglecting the factor p_1 leads to significant overestimation of the detection probability, with serious consequences on the reliability of CR systems that must meet tight requirements in terms of missed-detection rate. For this reason, it is advisable to design the decision threshold considering the worst-case condition in terms of p_1 and ρ, so as to ensure a missed-detection probability always lower than the maximum allowed level.

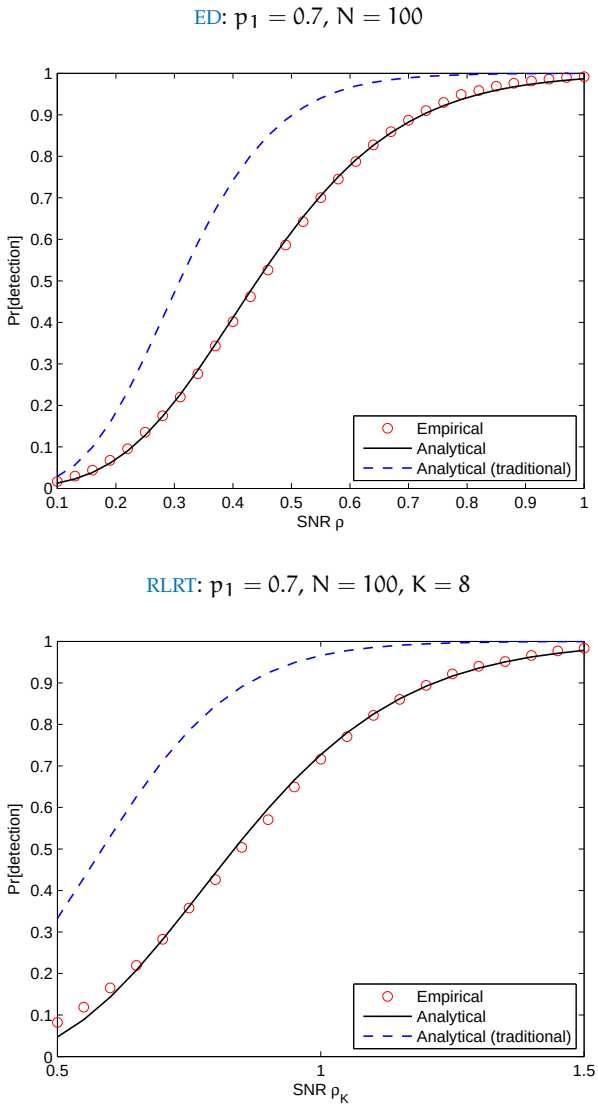
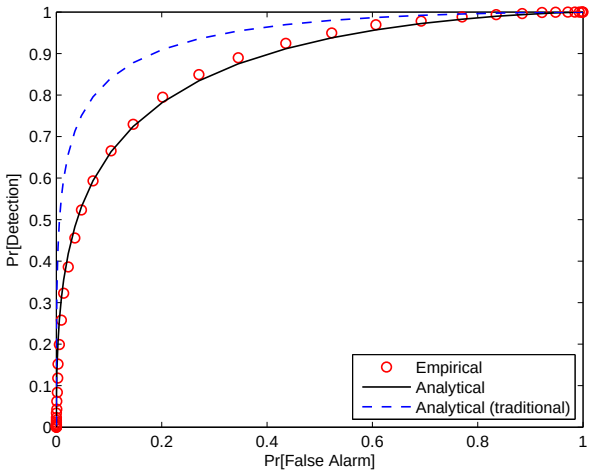


Figure 4.6: P_d vs. SNR for ED and RLRT, with fixed false-alarm rate $P_{fa} = 0.1\%$. Case of Gaussian signals.

ED: $p_1 = 0.7$, $N = 100$, $\rho = -6$dB



RLRT: $p_1 = 0.7$, $N = 50$, $K = 8$, $\rho_K = 0$dB

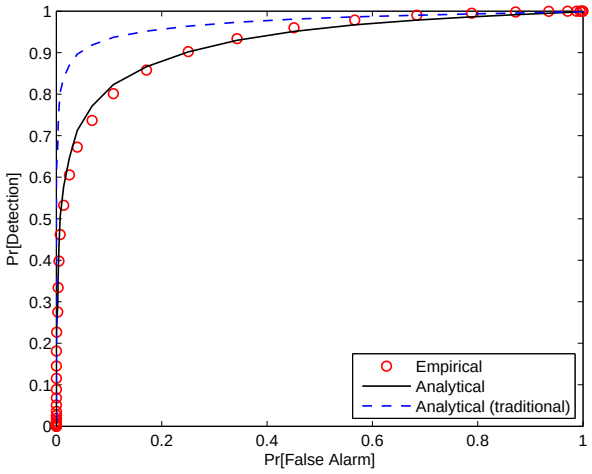


Figure 4.7: ROC for the considered methods. Case of Gaussian signals.

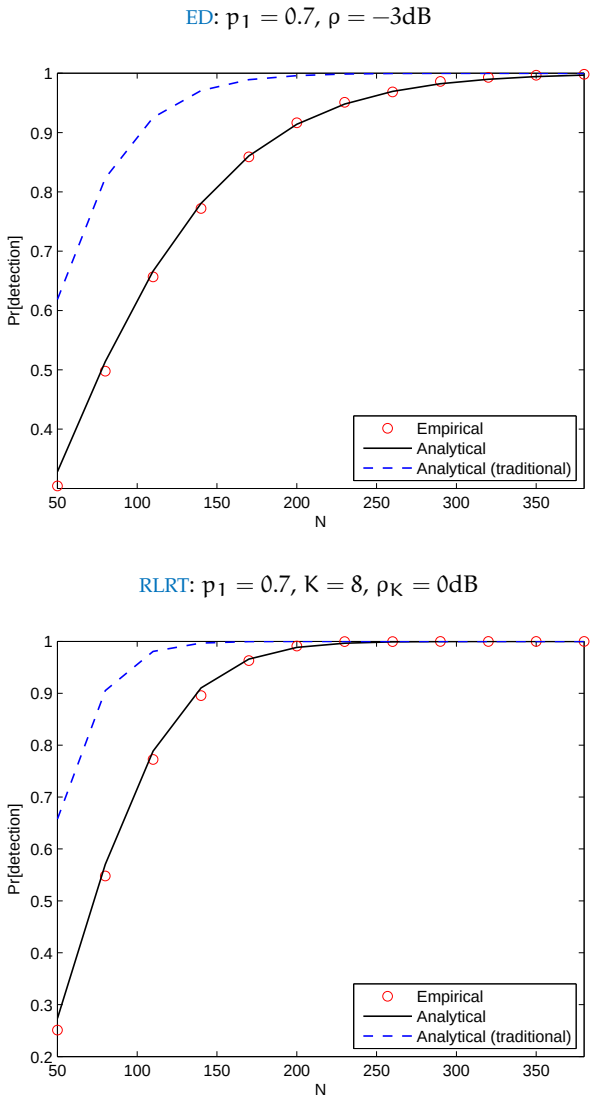


Figure 4.8: P_d vs. N for ED and RLRT, with fixed false-alarm rate $P_{fa} = 0.1\%$. Case of Gaussian signals.

Part III

BELIEF PROPAGATION FOR DISTRIBUTED INFERENCE

5

PROBABILISTIC GRAPHICAL MODELS AND DISTRIBUTED INFERENCE

5.1 INFERENCE THROUGH MESSAGE PASSING

5.1.1 *Bayesian Inference*

Several problems in wireless networks can be formulated as *Bayesian inference problems*. In general, consider a vector of N **hidden variables**

$$\mathbf{x} = [x_1, x_2, \ldots, x_N] \tag{5.1}$$

and a (fixed) vector of **observations** $\boldsymbol{y}$. The goal of a Bayesian inference procedure is to compute the *marginal a posteriori probabilities* for each of the variables of interest, i.e.,

$$p(x_n|\boldsymbol{y}) = \sum_{\boldsymbol{x}\setminus\{x_n\}} p(\mathbf{x}|\mathbf{y}). \tag{5.2}$$

of each of the hidden variables. In the above expression, notation $\boldsymbol{x}\setminus\{x_n\}$ indicates all elements of $\boldsymbol{x}$ except x_n, and summation is performed over the support of every variable involved. Discrete variables are assumed; in the case of continuous variables, the summation should be simply replaced by an integral, as follows:

$$p(x_n|\boldsymbol{y}) = \int p(\mathbf{x}|\mathbf{y}) d\{\boldsymbol{x}\setminus\{x_n\}\}, \tag{5.3}$$

where $d\{\boldsymbol{x}\setminus\{x_n\}\}$ denotes integration with respect to all elements of $\boldsymbol{x}$ except x_n over their respective supports.

Function $p(\mathbf{x}|\mathbf{y})$ is the *joint a posteriori probability* and, by using the well-known Bayes rule, it can be written as

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{5.4}$$

Bayesian inference problems can be conveniently represented in a graphical form by using *probabilistic graphical models*, as will be discussed next.

5.1.2 *Factor Graphs and Belief Propagation*

Direct computation of $p(x_n|\mathbf{y})$ is often intractable, due to the high-dimensional nature of $\mathbf{x}$. Practical algorithms can be developed relying on the factorization of both $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$. The factorization depends on conditional independence of the variables and is problem-specific.

Once factorized, the posterior probability $p(\mathbf{x}|\mathbf{y})$ can be conveniently represented through a **factor graph** (FG), i.e., a bipartite graph where each variable is represented by a *variable node* and each factor of the probability function as a *factor node*. By convention, variable nodes are typically depicted as circles, factor nodes as squares. Edges connect each factor node to all variables nodes corresponding to the arguments of the factor.

FGs allow for an efficient computation of the marginal probabilities of interest, by applying belief propagation (BP) and specifically the sum-product algorithm (SPA). This iterative procedure is based on message passing between factor and variable nodes in the graph, and produces as a result *beliefs* for every variable node, updated at each iteration. Beliefs are approximations of the marginal probabilities of interest. It is well-known that, if the FG is cycle-free [62] or contains one cycle at most [152], beliefs converge to the marginal a posteriori probabilities, hence BP is said to be *exact*. Otherwise, convergence of beliefs is not guaranteed [85]. In this case, the BP algorithm is sometimes referred to as "loopy" belief propagation (LBP). Although there is no theoretical guarantee of convergence, loopy belief propagation (LBP) is still successfully used as an approximate inference tool in many practical applications, the most famous of which being low-density parity-check (LDPC) and turbo decoding [76, 153].

Message update rules are expressed as follows. Denote with $f(\boldsymbol{x}_f)$ a generic factor connected to a subset of variables $\boldsymbol{x}_f \subset \boldsymbol{x}$; consider then a variable $x_i \in \boldsymbol{x}_f$, connected to factors $\mathcal{F}$ (including f). Then, messages sent from variable node x_i to factor node f have the form

$$\mu_{x_i \to f}(x_i) \propto \prod_{g \in \mathcal{F} \setminus \{f\}} \mu_{g \to x_i}(x_i) \tag{5.5}$$

and messages from factor node f to variable node x_i are defined as

$$\mu_{f \to x_i}(x_i) \propto \sum_{\boldsymbol{x}_f \setminus \{x_i\}} f(\boldsymbol{x}_f) \prod_{x_j \in \boldsymbol{x}_f \setminus \{x_i\}} \mu_{x_j \to f}(x_j). \tag{5.6}$$

Thus, messages to and from variable node x_i are probability functions having x_i as argument. The proportionality sign in (5.5) and (5.6) indicates the existence of proportionality factor such that the distribution sums/integrates to 1. However, messages do not necessarily need to be normalized at every iteration. Note that the summation in (5.6) must be replaced by an integral in the case of continuous variables.

Messages of type (5.5) and (5.6) are exchanged iteratively by every factor and variable nodes in the graph. For any variable node x_i, beliefs are computed as the product of all incoming messages:

$$b(x_i) \propto \prod_{f \in \mathcal{F}} \mu_{f \to x_i}(x_i). \tag{5.7}$$

After a sufficient number of iterations, beliefs are used as approximations of the marginal probabilities $p(x_i|\boldsymbol{y})$.

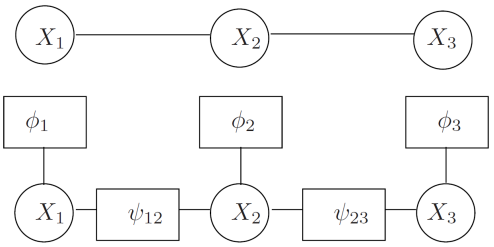


Figure 5.1: Markov random field (top) and factor graph (bottom) of the factorization $p(\mathbf{x}|\mathbf{y}) \propto \phi_1(x_1)\phi_2(x_2)\phi_3(x_3)\psi_{12}(x_1, x_2)\psi_{23}(x_2, x_3)$.

At the first iteration, all messages can be initialized as uniform distribution functions. Message scheduling (e.g., whether to update first messages from factors or messages from variables) has no impact on final results in the non-loopy case. On the contrary, scheduling may sometimes affect results when cycles are present.

5.1.3 *Pairwise Factors: Markov Random Fields*

Several problems of practical interest admit a factorization of the form

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^{N} \phi_n(x_n) \prod_{l=1}^{L} \psi_l(x_{l_1}, x_{l_2}), \tag{5.8}$$

where two types of factors exist:

- *single-variable factors* ϕ_n, representing individual prior distributions of variables x_n, and
- *pairwise factors* ψ_l, representing an *a priori* inter-dependency between variables x_{l_1} and x_{l_2}.

In this case, the FG model can be simplified by representing only variables as vertices of a graph, and pairwise factors as edges. Single-variables need not be depicted explicitly as they are connected to each variable. Such a model is referred to as **Markov random field** (MRF). Fig. 5.1 illustrates FG and MRF models for a simple factorization with three variables and two pairwise factors. Note that sometimes the short-hand notation $\psi_{ij}(x_i, x_j)$ is used instead of $\psi_l(x_{l_1}, x_{l_2})$.

When adopting a pairwise MRF representation, the message passing rules (5.5) and (5.6) can be expressed directly in a variable-to-variable form. Thus,

given two neighboring nodes x_i and x_j, and denoting by $\mathcal{N}_i$ the set of neighboring nodes of x_i, the message from x_i to x_j is

$$\mu_{x_i \to x_j}(x_j) \propto \sum_{x_i} \left(\phi_i(x_i)\psi_{ij}(x_i, x_j) \prod_{x_n \in \mathcal{N}_i \setminus \{x_j\}} \mu_{x_n \to x_i}(x_i) \right). \tag{5.9}$$

Beliefs are updated as

$$b_i(x_i) \propto \phi_i(x_i) \prod_{x_n \in \mathcal{N}_i} \mu_{x_n \to x_i}(x_i). \tag{5.10}$$

5.2 APPLICATION TO WIRELESS NETWORKS

5.2.1 *Network Message Passing*

The BP algorithm, run over FG or MRF models, is a powerful tool to address efficiently inference problems involving a large number of variables. Thus, probabilistic graphical models are constructed by representing hidden variables of the problem as variable nodes, and statistical dependencies among those variables as factors (FG) or edges (MRF).

In the case of *heterogeneous detection/estimation* problems in wireless networks, each hidden variable is typically associated to a node of the network, so that a one-to-one correspondence between vertices of the graph (variable nodes) and physical network nodes can be established. If statistical dependencies among hidden variables are pairwise, messages exchanged according to the BP algorithm (5.9) can be implemented as real messages (i.e., packets) exchanged over the wireless network. This approach, called *network message passing* [158], gives rise to a fully *decentralized* implementation of BP and can be applied to a variety of problems. In other problems, when multiple hidden variables exist but cannot be directly mapped to physical nodes, BP can be still adopted as an efficient tool to perform Bayesian inference in a wireless network, although adopting a *centralized* implementation. In the centralized case, BP messages are updated internally by the fusion center using the information coming from multiple, distributed sensor nodes.

5.2.2 *Examples*

Three examples of application of BP in wireless network are considered in this thesis.

1. **Hybrid cooperative localization.** Variables to be estimated are positions of nodes in a wireless network; observations are range (peer-to-peer distance) and pseudorange (distance from satellite) measurements. The problem is modeled by pairwise factors accounting for range measurements and individual factors as likelihood functions of

pseudorange measurement. Such a model makes it possible to implement BP as a network message passing algorithm. Thus, hybrid cooperative localization is formulated as a *heterogeneous, decentralized estimation problem* with *continuous, multi-dimensional hidden variables* (2-D or 3-D position vectors).

2. **Multi-sensor signal detection in heterogeneous environments.** The problem involves binary hidden variables representing signal presence/absence at different sensor locations. As a graphical model, a MRF representation can be adopted: individual factors account for the likelihood of nodes's noisy observations (received signal samples), and pairwise factors model statistical interdependency between the state of each node and that of its neighbors (as a result of the unknown spatial location of signal sources). Thus, the problem is formulated as a *heterogeneous, decentralized detection problem* with *discrete hidden variables*.

3. **Multi-sensor signal detection with malicious sensor nodes.** The considered scenario is a cognitive network performing cooperative spectrum sensing, such that distributed sensors send a binary decision to a fusion center. It is assumed that some of the sensors may be malicious and deliberately send wrong sensing reports with a certain "attack probability". Thus, two types of variables need to be estimated by the fusion center: a global, binary variable indicating signal presence (common to all sensors), and K variables indicating the type and probability of attack of each of the K sensors. Such variables are modeled as continuous variables in the range $[-1, 1]$: the sign indicates the type of attacker, the absolute value represents the attack probability. In this case, BP is run by the fusion center to estimate all hidden variables of interest. The inference problem is formulated as a *centralized* detection/estimation problem, involving a *mixture of homogeneous and heterogeneous* hidden variables.

The above applications are presented in details in the subsequent Chapters (6, 7, 8) of this thesis. Then, a variation of the BP algorithm will be introduced in Chapter 9.

6 HYBRID COOPERATIVE LOCALIZATION

6.1 MATHEMATICAL FORMULATION

Consider a network composed by a set of agents $\mathcal{M}$ of cardinality M and a set of satellites $\mathcal{S}$ of cardinality S. Time is slotted $t_0, t_1, \ldots, t_k$ and will be simply denoted by the discrete time index k. Referring to a particular agent $m \in \mathcal{M}$, denote by $\mathcal{M}_m^{(k)}$ the neighbors, i.e., the subset of peers agent m can communicate with at time k, and by $\mathcal{S}_m^{(k)}$ the subset of satellites it can see. Position variables are denoted by $\boldsymbol{p}_i^{(k)} = [x_i^{(k)}\ y_i^{(k)}\ z_i^{(k)}]^T$, where x, y, z are Earth-centered, Earth-fixed (ECEF) coordinates of agent/satellite i, and the superscript T denotes transpose. The clock bias of node m is denoted by $\delta_m^{(k)}$ and can be converted into distance units through the relation $b_m^{(k)} = c \cdot \delta_m^{(k)}$, where c is the speed of light. Thus, the state of node m is defined as

$$\boldsymbol{x}_m^{(k)} \triangleq \left[\boldsymbol{p}_m^{(k)T}\ b_m^{(k)}\right]^T. \tag{6.1}$$

In the considered hybrid scenario, two types of measurements are performed by nodes:

(i) Pseudoranges: estimated distances from satellites

$$\rho_{s\to m}^{(k)} = \|\boldsymbol{p}_s^{(k)} - \boldsymbol{p}_m^{(k)}\| + b_m^{(k)} + \nu_{s\to m}^{(k)}, \tag{6.2}$$

(ii) Ranges: estimated distances between peers

$$r_{n\to m}^{(k)} = \|\boldsymbol{p}_n^{(k)} - \boldsymbol{p}_m^{(k)}\| + \nu_{n\to m}^{(k)}, \tag{6.3}$$

where the symbol $\|\cdot\|$ denotes Euclidean distance, $m, n \in \mathcal{M}$, $s \in \mathcal{S}$, $\nu_{s\to m}^{(k)}$ and $\nu_{n\to m}^{(k)}$ are AWGN variables.

Notice that pseudorange measurements are affected by the additional unknown $b_m^{(k)}$, that is one of the variables to be estimated.[1] The following set notation is introduced to group together different nodes' variables: $\mathbb{x}_{\mathcal{M}}^{(k)} \triangleq \{\boldsymbol{x}_m^{(k)} | \forall m \in \mathcal{M}\}$; $\boldsymbol{\rho}_m^{(k)} \triangleq \{\rho_{s\to m}^{(k)} | \forall s \in \mathcal{S}_m^{(k)}\}$; $\mathbb{p}_{\mathcal{M}}^{(k)} \triangleq \{\boldsymbol{\rho}_m^{(k)} | \forall m \in \mathcal{M}\}$; $\boldsymbol{r}_m^{(k)} \triangleq \{r_{n\to m}^{(k)} | \forall n \in \mathcal{M}_m^{(k)}\}$; $\mathbb{r}_{\mathcal{M}}^{(k)} \triangleq \{\boldsymbol{r}_m^{(k)} | \forall m \in \mathcal{M}\}$.

The localization problem can be formulated as follows: every agent m needs to determine its *a posteriori* distribution of $\boldsymbol{x}_m^{(k)}$, at each time slot k, given all the available measurements:

$$p\left(\boldsymbol{x}_m^{(k)} \middle| \mathbb{p}_{\mathcal{M}}^{(1:k)}, \mathbb{r}_{\mathcal{M}}^{(1:k)}\right), \quad \forall m \in \mathcal{M}, \tag{6.4}$$

1 Peer-to-peer range measurements can be performed by methods that avoid synchronization, like round-trip-time [32] or received signal strength measurements [42].

where the superscript $(1:k)$ denotes all variables from time step 1 to k, e.g. $\mathbb{r}_{\mathcal{M}}^{(1:k)} \triangleq \{\mathbb{r}_{\mathcal{M}}^{(1)}, \dots, \mathbb{r}_{\mathcal{M}}^{(k)}\}$.

The following assumptions, which are reasonable approximations in many practical scenarios, are made in this section:

A1: Mobility is modeled as a Markov process, mutually independent for every node:[2]

$$p\left(\mathbb{x}_{\mathcal{M}}^{(k)} \middle| \mathbb{x}_{\mathcal{M}}^{(k-1)}, \cdots, \mathbb{x}_{\mathcal{M}}^{(0)}\right) = \prod_{m \in \mathcal{M}} p\left(\boldsymbol{x}_m^{(k)} \middle| \boldsymbol{x}_m^{(k-1)}\right). \tag{6.5}$$

A2: Measurement likelihood depends only on the current state and can be split into two factors, since range and pseudorange measurements are independent:

$$p\left(\mathbb{p}_{\mathcal{M}}^{(k)}, \mathbb{r}_{\mathcal{M}}^{(k)} \middle| \mathbb{x}_{\mathcal{M}}^{(0:k)}\right) = p\left(\mathbb{p}_{\mathcal{M}}^{(k)} \middle| \mathbb{x}_{\mathcal{M}}^{(k)}\right) p\left(\mathbb{r}_{\mathcal{M}}^{(k)} \middle| \mathbb{x}_{\mathcal{M}}^{(k)}\right). \tag{6.6}$$

A3: Pseudorange measurement noise samples are independent with variance known by each node:

$$\nu_{s\to m}^{(k)} \sim \mathcal{N}\left(0, \sigma_{s\to m}^2\right). \tag{6.7}$$

A4: Range measurement noise samples are independent, with symmetric link variance and known by both nodes:

$$\nu_{m\to n}^{(k)}, \nu_{n\to m}^{(k)} \sim \mathcal{N}\left(0, \sigma_{m\to n}^2\right). \tag{6.8}$$

6.2 CRAMÉR-RAO BOUND

The CRB of any unbiased estimator of $[\boldsymbol{X}, \boldsymbol{b}]$ is obtained by inverting the corresponding Fisher information matrix (FIM). Let $\boldsymbol{F}$ be the FIM for the hybrid scenario. First, the FIM is computed under a non-cooperative setting, and then extend this result to the cooperative case.

6.2.1 *Non-cooperative Case*

Focus on a single agent, say m. The log-likelihood function of its measurements with respect to anchors and satellites is

$$\begin{aligned}
&\log p\left(\{r_{a\to m}\}_{a\in\mathcal{A}_m}, \{\rho_{s\to m}\}_{s\in\mathcal{S}_m} | \boldsymbol{x}_m, b_m\right) \\
&= \sum_{a\in\mathcal{A}_m} \log p\left(r_{a\to m} | \boldsymbol{x}_m\right) + \sum_{s\in\mathcal{S}_m} \log p\left(\rho_{s\to m} | \boldsymbol{x}_m, b_m\right) \\
&\doteq \Lambda_m\left(\boldsymbol{x}_m, b_m\right).
\end{aligned}$$

2 It is assumed that each node has some internal information about its own mobility model, expressed by a PDF $p\left(\boldsymbol{x}_m^{(k)} \middle| \boldsymbol{x}_m^{(k-1)}\right)$. However, this distribution may differ from the true mobility statistics, which depend on the users in the network. See for example simulations in Sec. 6.5.2.2.

Under Gaussian measurement noise:

$$\log p\left(r_{a\to m}\,|\boldsymbol{x}_m\right) = C - \frac{|r_{a\to m} - \|\boldsymbol{x}_a - \boldsymbol{x}_m\||^2}{2\sigma^2_{a\to m}}$$

and

$$\log p\left(\rho_{s\to m}\,|\boldsymbol{x}_m, b_m\right) = C' - \frac{|\rho_{s\to m} - \|\boldsymbol{x}_s - \boldsymbol{x}_m\| - b_m|^2}{2\sigma^2_{s\to m}},$$

where C, C' are constant terms. The Fisher information matrix is given by

$$\boldsymbol{F}_m = -\mathbb{E}\{\,H_m\left(\Lambda_m\left(\boldsymbol{x}_m, b_m\right)\right)\},$$

where the expectation is with respect to the measurements, and $H_m(\cdot)$ is the Hessian operator containing the second-order partial derivatives with respect to each element of $[\boldsymbol{x}_m, b_m]$. $\boldsymbol{F}_m$ is a $(D+1)\times(D+1)$ matrix:

$$\boldsymbol{F}_m = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{x}_m} & \boldsymbol{f}_{\boldsymbol{x}_m, b_m} \\ \boldsymbol{f}^T_{\boldsymbol{x}_m, b_m} & F_{b_m} \end{bmatrix} \succeq 0, \tag{6.9}$$

where

$$\begin{aligned} \boldsymbol{F}_{\boldsymbol{x}_m} &= \sum_{a\in\mathcal{A}_m} \frac{1}{\sigma^2_{a\to m}} \boldsymbol{q}_{am}\boldsymbol{q}^T_{am} + \sum_{s\in\mathcal{S}_m} \frac{1}{\sigma^2_{s\to m}} \boldsymbol{q}_{sm}\boldsymbol{q}^T_{sm} \\ F_{b_m} &= \sum_{s\in\mathcal{S}_m} \frac{1}{\sigma^2_{s\to m}} \\ \boldsymbol{f}_{\boldsymbol{x}_m, b_m} &= \sum_{s\in\mathcal{S}_m} -\frac{1}{\sigma^2_{s\to m}} \boldsymbol{q}_{sm}, \end{aligned}$$

in which $\boldsymbol{q}_{im} = \frac{\boldsymbol{x}_i - \boldsymbol{x}_m}{\|\boldsymbol{x}_i - \boldsymbol{x}_m\|}$ is a unit-length column vector between $\boldsymbol{x}_m$ and $\boldsymbol{x}_i$.

Considering all M agents, the global non-cooperative FIM is a block-diagonal matrix:

$$\boldsymbol{F}_{\text{non-coop}} = \begin{bmatrix} \boldsymbol{F}_1 & & & \\ & \boldsymbol{F}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{F}_M \end{bmatrix}. \tag{6.10}$$

6.2.2 *Cooperative Case*

The log-likelihood function is now

$$\begin{aligned} &\log p\left(\left\{\{r_{a\to m}\}_{a\in\mathcal{A}_m}, \{\rho_{s\to m}\}_{s\in\mathcal{S}_m}, \right.\right. \\ &\quad \left.\left.\{r_{n\to m}\}_{n\in\mathcal{M}_m}\right\}_{m\in\mathcal{M}} |\boldsymbol{X}, \boldsymbol{b}\right) \\ &= \sum_{m\in\mathcal{M}} \Lambda_m\left(\boldsymbol{x}_m, b_m\right) + \underbrace{\sum_{m\in\mathcal{M}} \sum_{n\in\mathcal{M}_m} \log p\left(r_{n\to m}\,|\boldsymbol{x}_m, \boldsymbol{x}_n\right)}_{\doteq \Lambda_{\text{coop}}(\boldsymbol{X})}. \end{aligned}$$

The Fisher information matrix is of the form

$$\boldsymbol{F} = \boldsymbol{F}_{\text{non-coop}} + \boldsymbol{F}_{\text{coop}} \tag{6.11}$$

and has dimension $(D+1)M \times (D+1)M$. The first term $\boldsymbol{F}_{\text{non-coop}}$, representing the non-cooperative contribution, is again (6.10). The cooperative part $\boldsymbol{F}_{\text{coop}}$ can be expressed as

$$\boldsymbol{F}_{\text{coop}} = -\mathbb{E}\left\{ \begin{bmatrix} H_{11} & \dots & H_{M1} \\ \vdots & \ddots & \vdots \\ H_{M1} & \dots & H_{MM} \end{bmatrix} \Lambda_{\text{coop}}(\boldsymbol{X}) \right\}$$

where the cross-Hessian matrices H_{mn} are defined as (assuming $\boldsymbol{x}_i = [x_{1,i}, \dots, x_{D,i}]$):

$$H_{mn} \doteq \begin{bmatrix} \frac{\partial^2}{\partial x_{1,m} \partial x_{1,n}} & \dots & \frac{\partial^2}{\partial x_{1,m} \partial x_{D,n}} & \frac{\partial^2}{\partial x_{1,m} \partial b_n} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2}{\partial x_{D,m} \partial x_{1,n}} & \dots & \frac{\partial^2}{\partial x_{D,m} \partial x_{D,n}} & \frac{\partial^2}{\partial x_{D,m} \partial b_n} \\ \frac{\partial^2}{\partial b_m \partial x_{1,n}} & \dots & \frac{\partial^2}{\partial b_m \partial x_{D,n}} & \frac{\partial^2}{\partial b_m \partial b_n} \end{bmatrix}$$

Notice that $\Lambda_{\text{coop}}(\boldsymbol{X})$ does not depend on the bias. Under the hypothesis of Gaussian measurement noise in peer-to-peer communication,

$$\log p(r_{n\to m} | \boldsymbol{x}_m, \boldsymbol{x}_n) = C'' - \frac{|r_{n\to m} - \|\boldsymbol{x}_n - \boldsymbol{x}_m\||^2}{2\sigma^2_{n\to m}},$$

leading to a block matrix of the form

$$\boldsymbol{F}_{\text{coop}} = \begin{bmatrix} \boldsymbol{F}'_1 & \mathbf{o} & \boldsymbol{K}_{12} & \mathbf{o} & \dots & \boldsymbol{K}_{1M} & \mathbf{o} \\ \mathbf{o}^T & 0 & \mathbf{o}^T & 0 & & \mathbf{o}^T & 0 \\ \boldsymbol{K}_{21} & \mathbf{o} & \boldsymbol{F}'_2 & \mathbf{o} & & & \\ \mathbf{o}^T & 0 & \mathbf{o}^T & 0 & & & \\ \vdots & & & & \ddots & & \\ \boldsymbol{K}_{M1} & \mathbf{o} & & & & \boldsymbol{F}'_M & \mathbf{o} \\ \mathbf{o}^T & 0 & & & & \mathbf{o}^T & 0 \end{bmatrix} \succeq 0. \tag{6.12}$$

where

$$\begin{aligned} \boldsymbol{F}'_m &= \sum_{n \in \mathcal{M}_m} \frac{1}{\sigma^2_{n\to m}} \boldsymbol{q}_{nm} \boldsymbol{q}_{nm}^T \\ \boldsymbol{K}_{mn} &= \begin{cases} -\frac{1}{\sigma^2_{n\to m}} \boldsymbol{q}_{nm} \boldsymbol{q}_{nm}^T, & \text{if } n \in \mathcal{M}_m \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

and $\mathbf{o}$ is a $D \times 1$ zero-vector.

The above results allow to compute $\boldsymbol{F}$ for a given network configuration and, by inverting (6.11), to express the CRB.

6.3 BAYESIAN INFERENCE ON FACTOR GRAPH

The goal of a Bayesian positioning approach is to determine the marginals (6.4) recursively at each time slot k. Assume to be given a prior distribution $p(\boldsymbol{x}_m^{(0)})$, $\forall m \in \mathcal{M}$, which may be uniform if no *a priori* information is available. Marginals are given by

$$p\left(\boldsymbol{x}_m^{(k)} \middle| \mathbb{p}_{\mathcal{M}}^{(1:k)}, \mathbb{r}_{\mathcal{M}}^{(1:k)}\right) = \int p\left(\mathbb{x}_{\mathcal{M}}^{(k-1:k)} \middle| \mathbb{p}_{\mathcal{M}}^{(1:k)}, \mathbb{r}_{\mathcal{M}}^{(1:k)}\right) d\mathbb{x}_{\mathcal{M}\setminus m}^{(k-1:k)}, \tag{6.13}$$

where the subscript $\mathcal{M}\setminus m$ denotes all variables in $\mathbb{x}_{\mathcal{M}}^{(k-1:k)}$ except $\boldsymbol{x}_m^{(k)}$. By **A1-A4**, the joint *a posteriori* distribution becomes

$$\begin{aligned} p\left(\mathbb{x}_{\mathcal{M}}^{(k-1:k)} \middle| \mathbb{p}_{\mathcal{M}}^{(1:k)}, \mathbb{r}_{\mathcal{M}}^{(1:k)}\right) &= p\left(\mathbb{p}_{\mathcal{M}}^{(k)} \middle| \mathbb{x}_{\mathcal{M}}^{(k)}\right) p\left(\mathbb{r}_{\mathcal{M}}^{(k)} \middle| \mathbb{x}_{\mathcal{M}}^{(k)}\right) \times \\ &\prod_{m\in\mathcal{M}} p\left(\boldsymbol{x}_m^{(k)} \middle| \boldsymbol{x}_m^{(k-1)}\right) p\left(\boldsymbol{x}_m^{(k-1)} \middle| \mathbb{p}_{\mathcal{M}}^{(1:k-1)}, \mathbb{r}_{\mathcal{M}}^{(1:k-1)}\right). \end{aligned} \tag{6.14}$$

The first two factors represent the likelihood of range and pseudorange measurements, respectively, while the subsequent factors account for temporal evolution of each node's state variable according to a given mobility model. The last factor in (6.14) is the *a posteriori* distribution of agent m at the previous time slot. Hence, the *a posteriori* distributions can be computed recursively. Written in this form, the above probability function can be mapped on a FG, which allows to compute efficiently the marginals (6.4) by applying the SPA [62]. A FG is a probabilistic graphical model that represents statistical dependencies among variables, and is used to perform Bayesian inference. Due to GNSS bias, there are two options to construct the FG: *(a)* considering position and bias variables jointly, or *(b)* treating them separately. Both options are next analyzed and discussed.

6.3.1 *Joint Position-Bias Model*

Given as input

$$p(\boldsymbol{x}_m^{(k-1)} | \mathbb{p}_{\mathcal{M}}^{(1:k-1)}, \mathbb{r}_{\mathcal{M}}^{(1:k-1)}) \quad \forall m, \tag{6.15}$$

i.e., the result of inference process at the previous time slot, the remaining part of (6.14) can be decomposed as

$$\prod_{m=1}^{M} \left[f_m\left(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_m^{(k-1)}\right) \prod_{s\in\mathcal{S}_m} g_{s,m}\left(\boldsymbol{x}_m^{(k)}\right) \prod_{\substack{n\in\mathcal{M}_m \\ n<m}} h_{n,m}\left(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_n^{(k)}\right) \right], \tag{6.16}$$

where

- $f_m(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_m^{(k-1)}) \equiv p(\boldsymbol{x}_m^{(k)} | \boldsymbol{x}_m^{(k-1)})$ represents mobility.

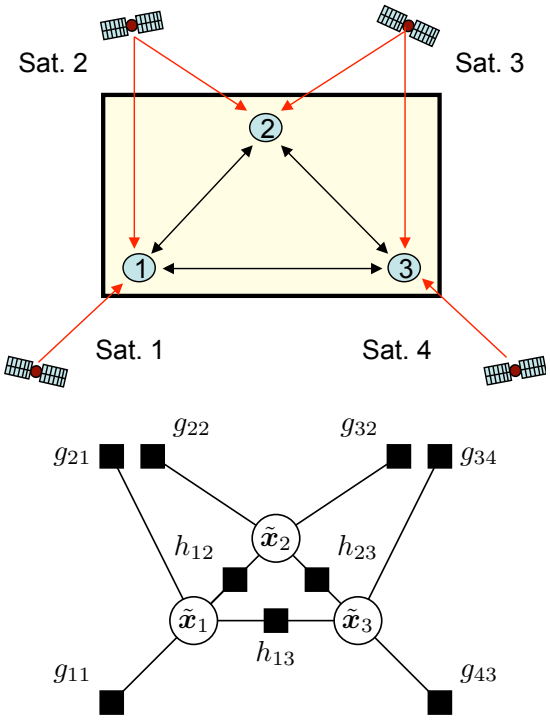


Figure 6.1: Example of GNSS-cooperative network and corresponding factor graph (joint position-bias variable representation).

- $g_{s,m}(\boldsymbol{x}_m^{(k)}) \equiv p(\rho_{s\to m}|\boldsymbol{x}_m^{(k)})$ represents the pseudorange measurement likelihood given the state of node m. Note that positions of the satellites do not appear as variables, since they are known exactly by nodes.[3]
- $h_{n,m}(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_n^{(k)}) \equiv p(r_{n\to m}|\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_n^{(k)})$ represents the peer-to-peer range measurement likelihood given the positions of nodes m and n.

The resulting FG is depicted in Fig. 6.2(a), where the vertices on top receive messages $p(\boldsymbol{x}_m^{(k-1)}|\mathbb{p}_{\mathcal{M}}^{(1:k-1)}, \mathbb{r}_{\mathcal{M}}^{(1:k-1)})$ as inputs from the previous time slot.

Factors corresponding to the same node in the physical network are grouped into dashed boxes. Therefore, vertices of the FG, i.e., variables to be estimated, have been associated to nodes in the network, i.e., devices that perform computations and send messages to neighbors. Then, probabilistic messages defined by the SPA algorithm (see next section) correspond to real packets exchanged among physical nodes: in this way the resulting algorithm is fully distributed. Fig. 6.1 shows a network configuration example and its corresponding FG.

3 Satellite positions are computed through the *ephemeris* (i.e., orbital information) that is part of every message sent by GPS satellites.

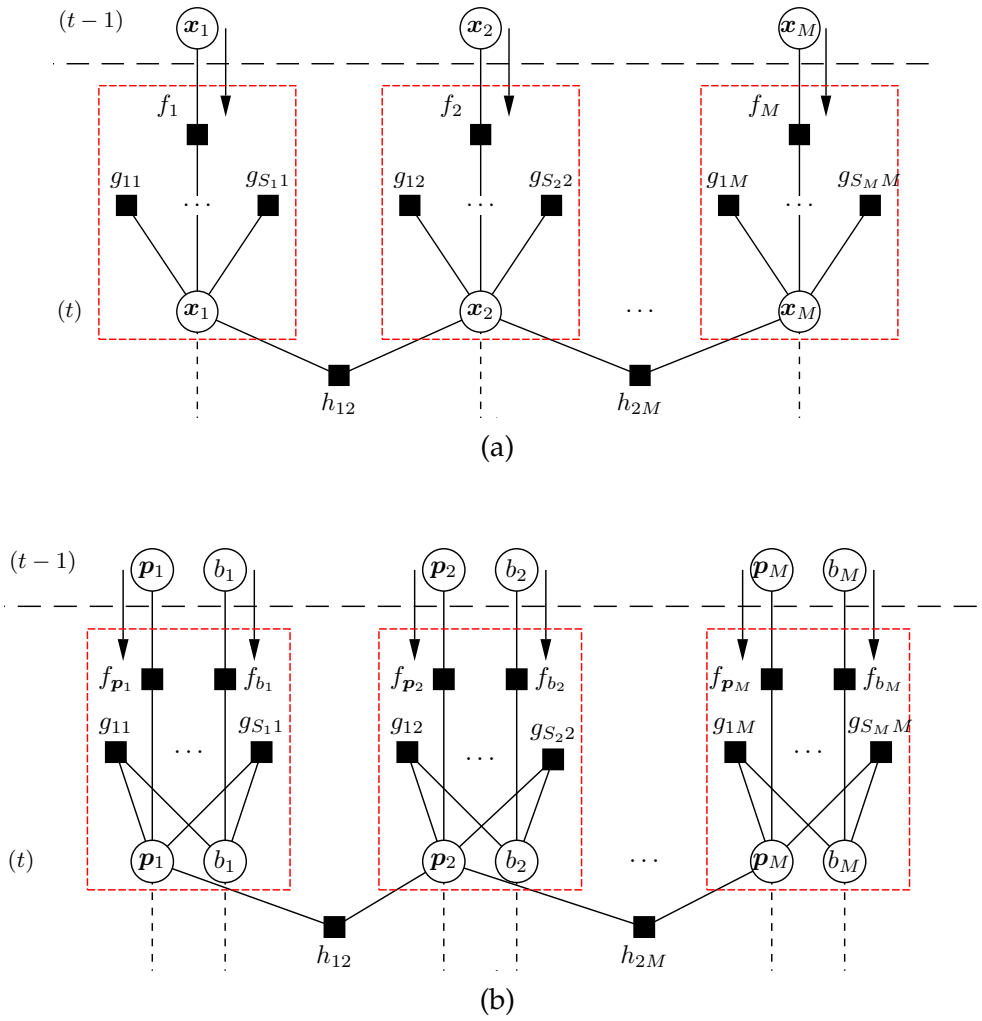


Figure 6.2: Factor graphs for hybrid cooperative positioning: (a) joint position-bias model, (b) separate position-bias model. Dashed boxes represent physical nodes, i.e., messages inside a box are computed internally by a node.

6.3.2 *Separate Position-Bias Model*

In the alternative model, separate temporal factors exist for position and bias; range factors involve position variables only, while pseudorange factors are connected to both position and bias variables. Given as input

$$p(\boldsymbol{x}_m^{(k-1)}|\mathbb{p}_{\mathcal{M}}^{(1:k-1)},\mathbb{r}_{\mathcal{M}}^{(1:k-1)}) \quad \forall m, \tag{6.17}$$

the factorization of the remaining part of (6.14) is

$$\prod_{m=1}^{M}\left[f_{\boldsymbol{p}_m}\left(\boldsymbol{p}_m^{(k)},\boldsymbol{p}_m^{(k-1)}\right) f_{b_m}\left(b_m^{(k)},b_m^{(k-1)}\right)\times \prod_{s\in\mathcal{S}_m} g_{s,m}\left(\boldsymbol{p}_m^{(k)},b_m^{(k)}\right)\prod_{\substack{n\in\mathcal{M}_m\\ n<m}} h_{n,m}\left(\boldsymbol{p}_m^{(k)},\boldsymbol{p}_n^{(k)}\right)\right]. \tag{6.18}$$

This model is represented by the FG in Fig. 6.2(b). Information from the previous time slot is passed by messages $p(\boldsymbol{p}_m^{(k-1)}|\mathbb{p}_{\mathcal{M}}^{(1:k-1)},\mathbb{r}_{\mathcal{M}}^{(1:k-1)})$ and $p(b_m^{(k-1)}|\mathbb{p}_{\mathcal{M}}^{(1:k-1)})$ sent by vertices at level $k-1$.

6.4 ALGORITHM DESCRIPTION

In this section the proposed H-SPAWN algorithm is presented. The section first presents a high-level description of the algorithm is provided, then focuses on message passing implementation over a wireless network. Then, message computation is discussed, introducing a parametric implementation strategy to efficiently update and exchange H-SPAWN messages.

6.4.1 *Summary of the Algorithm*

H-SPAWN is defined by applying SPA message update rules [62] over the factor graphs of Fig. 6.2a or 6.2b. Denoting by ξ_m a generic state variable (i.e., $\boldsymbol{x}_m$, $\boldsymbol{p}_m$, or b_m), messages can be classified as: (i) *temporal messages* $\eta_{f_m\to\xi_m}$, representing the evolution of state variable m from time $k-1$ to time k; (ii) *messages from satellite factors* $\eta_{g_{s,m}\to\xi_m}$, associated to GNSS measurements from satellites $s\in\mathcal{S}_m^{(k)}$; (iii) *messages from peer-to-peer factors* $\eta_{h_{n,m}\to\xi_m}$, representing the information received by node m from neighbors; (iv) *messages to peer-to-peer factors* $\eta_{\xi_m\to h_{mn}}$, whereby node m communicates its position message to neighbors $n\in\mathcal{M}_m^{(k)}$; (v) *messages to satellite factors* $\eta_{\boldsymbol{p}_m^{(k)}\to g_{s,m}}$ and $\eta_{b_m^{(k)}\to g_{s,m}}$, used only in the separate model to link bias and position variables together.

Thanks to the identity between variables and physical nodes, H-SPAWN admits a natural distributed implementation: messages of type (i), (ii), (iii), and (v) are computed internally by each node, whereas messages of type (iv) are sent in broadcast and received by all nodes that are close enough at

Algorithm 6.1: H-SPAWN – joint position-bias version

input : Initial beliefs $\hat{p}\left(\boldsymbol{x}_m^{(0)}\right)\ \forall m$

output: Updated beliefs $\hat{p}\left(\boldsymbol{x}_m^{(k)}\right)\ \forall m$

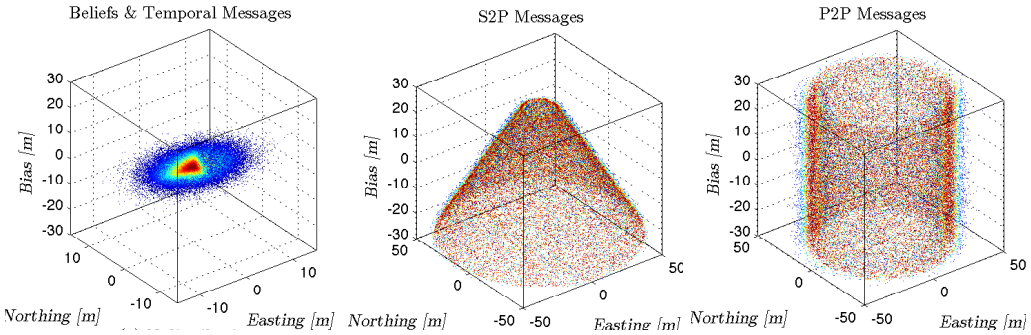
1 **for** *timestep* $k = 1$ *to* K **do**
2 $\quad\forall m \in \mathcal{M}$: compute temporal message $\eta_{f_m \to \boldsymbol{x}_m^{(k)}}$ using (6.19)
3 $\quad\forall m \in \mathcal{M}$: broadcast predicted distribution as initial peer-to-peer (P2P) message: $\eta_{\boldsymbol{x}_m^{(k)} \to h_{n,m}} = \eta_{f_m \to \boldsymbol{x}_m^{(k)}}$
4 $\quad\forall m \in \mathcal{M}$: collect ranges $r_{n\to m}\ \forall n \in \mathcal{M}_m^{(k)}$ and pseudoranges $\rho_{s\to m}$ $\forall s \in \mathcal{S}_m^{(k)}$
5 $\quad$**for** *iteration* $i = 1$ *to* I **do**
6 $\quad\quad$**for** *nodes* $m \in \mathcal{M}$ ***in parallel*** **do**
7 $\quad\quad\quad$Receive messages $\eta_{\boldsymbol{x}_n^{(k)} \to h_{n,m}}$ from all neighbors $n \in \mathcal{M}_m^{(k)}$
8 $\quad\quad\quad$Compute local factor-to-variable messages given the likelihood of received measurements from peers and from satellites: $\eta_{g_{s,m} \to \boldsymbol{x}_m^{(k)}}\ \forall s \in \mathcal{S}_m^{(k)}$; $\eta_{h_{n,m} \to \boldsymbol{x}_m^{(k)}}$ using (6.25) and (6.22)
9 $\quad\quad\quad$Compute and broadcast outgoing P2P messages $\eta_{\boldsymbol{x}_m^{(k)} \to h_{n,m}}$ $\forall n \in \mathcal{M}_m^{(k)}$ using (6.28)
10 $\quad\quad\quad$Update beliefs $\hat{p}\left(\boldsymbol{x}_m^{(k)}\right)$ using (6.30)
11 $\quad\quad$**end**
12 $\quad$**end**
13 **end**

time k (neighbor sets $\mathcal{M}_m$ may vary at different time slots). Beliefs $\hat{p}(\xi_m)$ (estimated marginal probabilities) are then computed as the product of all incoming messages towards ξ_m.

The resulting algorithm and a possible message scheduling are summarized by Alg. 6.1, referring to joint position-bias model. For the separate model, all messages involving variables $\boldsymbol{x}_m^{(k)}$ should be replaced by two separate messages for $\boldsymbol{p}_m^{(k)}$ and for $b_m^{(k)}$; in addition, between lines 8 and 9, messages $\eta_{\boldsymbol{p}_m^{(k)} \to g_{s,m}}$ and $\eta_{b_m^{(k)} \to g_{s,m}}\ \forall s \in \mathcal{S}_m^{(k)}$ must be updated. Observe that the algorithm works on two time scales: in the *iteration* scale, messages are updated using the same measurement data; in the *timestep* scale, new measurements are received.

6.4.2 *Parametric BP: Joint Position-Bias Model*

This section illustrates how the messages defined in the previous section are computed in practice. For each type of message, the exact update rules are expressed according to the SPA. Since these expressions are numerically complex (due to integrals and multiplications in 3-D or 4-D spaces), a *parametric BP* strategy can be adopted to avoid direct computation of messages. The parametric approach approximates all messages by known "distribution



(a) $\mathcal{N}$ distribution. (b) $\mathcal{V}$ distribution. (c) $\mathcal{C}$ distribution.

Figure 6.3: Distribution families for the joint position-bias model. Horizontal sections.

families". In this way, message passing in BP reduces to simply passing the *parameters* that characterize such distributions. The advantage of parametric BP over a sample-based message representation is enormous in terms of numerical complexity: the number of samples needed to represent messages would be prohibitive for any practical wireless network.

The FG in Fig. 6.2a is first considered, where position and bias variables are considered jointly; then, the separate case is analyzed in Sec. 6.4.3.

6.4.2.1 *Temporal messages*

The SPA update rule for temporal messages over the FG of Fig. 6.2a is

$$\eta_{f_m \to \boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right) \propto \int f_m\left(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_m^{(k-1)}\right) \times p\left(\boldsymbol{x}_m^{(k-1)} | \mathbb{p}_{\mathcal{M}}^{(1:k-1)}, \mathbb{r}_{\mathcal{M}}^{(1:k-1)}\right) d\boldsymbol{x}_m^{(k-1)}. \tag{6.19}$$

The above expression has the following meaning: each node propagates the beliefs of $\boldsymbol{x}_m = [\boldsymbol{p}_m^T, b_m]^T$ from time $k-1$ to time k, according to some internal mobility model (which updates the position $\boldsymbol{p}_m$) and clock drift model (which updates the bias b_m) jointly represented by factor f_m. Temporal messages can be interpreted as a *prediction step* in the localization process.

Adopting a parametric approach, temporal messages are approximated as $(D+1)$-variate Gaussian PDFs, with mean $\boldsymbol{\mu}_{\boldsymbol{x}_m^{(k)}}$ derived from $\boldsymbol{\mu}_{\boldsymbol{x}_m^{(k-1)}}$ using mobility and clock drift models, and covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k)}} \succeq \boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k-1)}}$ (e.g., linearly increasing with the time elapsed between t_{k-1} and t_k) which accounts for prediction uncertainty:

$$\eta_{f_m \to \boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right) \approx p_{\mathcal{N}}\left(\boldsymbol{x}_m^{(k)}; \boldsymbol{\mu}_{\boldsymbol{x}_m^{(k)}}, \boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k)}}\right), \tag{6.20}$$

with

$$p_{\mathcal{N}}(\boldsymbol{x}; \boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{x}}) \triangleq \frac{1}{Z} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})^T \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})\right], \tag{6.21}$$

where Z is a normalization factor that may be chosen arbitrarily (messages and beliefs in BP do not necessarily need to be normalized). The shape of the $\mathcal{N}$ distribution is a *hyper-ellipsoid*. A section of such distribution is depicted in Fig. 6.3a.

6.4.2.2 *Messages from satellite factors*

Due to the single connection of satellite factors, satellite messages simply propagate the function received from factor $g_{s,m}$:

$$\eta_{g_{s,m}\to\boldsymbol{x}_m^{(k)}} = g_{s,m}\left(\boldsymbol{x}_m^{(k)}\right), \tag{6.22}$$

i.e., the likelihood of measurement $\rho_{s\to m}^{(k)}$ with respect to the position of satellite s and node m. Messages of this type can be represented in a parametric form as

$$\eta_{g_{s,m}\to\boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right) = p_{\mathcal{V}}\left(\boldsymbol{x}_m^{(k)}; \rho_{s\to m}^{(k)}, \boldsymbol{p}_s^{(k)}, \sigma_{s\to m}^2\right), \tag{6.23}$$

where $\rho_{s\to m}^{(k)}$ is the value of the pseudorange (6.2) measured at the current time slot k, $\sigma_{s\to m}^2$ is the variance associated to this measurement (6.7), $\boldsymbol{p}_s^{(k)}$ is the satellite position, and

$$p_{\mathcal{V}}(\boldsymbol{x}; \rho, \boldsymbol{\mu}_s, \sigma_\rho^2) \triangleq \frac{1}{Z}\exp\left[-\frac{1}{2\sigma_\rho^2}\left(\|\boldsymbol{p}-\boldsymbol{\mu}_s\| + b - \rho\right)^2\right] \tag{6.24}$$

is a *hyper-conic* distribution family parametrized by radius ρ (pseudorange value), center $\boldsymbol{\mu}_s$ (satellite position) and variance σ_ρ^2 (measurement variance). Note that in this case parametric BP does not introduce any approximation.

The $\mathcal{V}$ distribution is in fact a Gaussian distribution having as support the surface of a sphere that shrinks or expands with the bias. Such support can be visualized as a spherical cone with spheres as intersections with hyper-planes perpendicular to the b-axis, and cones as intersections with $\boldsymbol{p}$ planes (hence the name of hyper-conic distribution). Fig. 6.3b shows a section of a $\mathcal{V}$ distribution.

6.4.2.3 *Messages from peer-to-peer factors*

The analytical expression of messages from P2P factors is

$$\eta_{h_{n,m}\to\boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right) \propto \int h_{n,m}\left(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_n^{(k)}\right) \times \eta_{\boldsymbol{x}_n^{(k)}\to h_{n,m}}\left(\boldsymbol{x}_n^{(k)}\right) d\boldsymbol{x}_n^{(k)}. \tag{6.25}$$

In (6.25), the term $h_{n,m}(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_n^{(k)})$ represents the likelihood of the range measurement $\rho_{n\to m}^{(k)}$ received by node m from its neighbor n, given the position $\boldsymbol{x}_n^{(k)}$; the message $\eta_{\boldsymbol{x}_n^{(k)}\to h_{n,m}}(\boldsymbol{x}_n^{(k)})$ provides a PDF of $\boldsymbol{x}_n^{(k)}$, based on

the available information at node n (i.e., measurements from visible satellites and all neighbors except m). Thus, the message update rule can be interpreted as a *filtering* operation of the range likelihood function with the estimated PDF of the position of node n.

In order to express P2P messages in a parametric form, let us first consider $\boldsymbol{x}_n^{(k)} = \hat{\boldsymbol{x}}_n^{(k)}$ fixed, as though $\eta_{\boldsymbol{x}_n^{(k)} \to h_{n,m}}(\boldsymbol{x}_n^{(k)})$ were a Dirac delta function in $\hat{\boldsymbol{x}}_n^{(k)}$. Then, the resulting message would amount to $h_{n,m}(\boldsymbol{x}_m^{(k)}, \hat{\boldsymbol{x}}_n^{(k)})$. This message can be visualized as a Gaussian distribution with variance $\sigma^2_{n\to m}$ around the surface of a *hyper-cylinder* in the $\{\boldsymbol{p}, b\}$ plane, uniformly extended over the entire range of b (because P2P messages do not carry bias information), with axis $\boldsymbol{p} = \hat{p}_n^{(k)}$ and radius $\rho_{n\to m}^{(k)}$. Now, filtering $h_{n,m}$ with $\eta_{\boldsymbol{x}_n^{(k)} \to h_{n,m}}$ is a summation of an infinite number of such cylinders with different axes. The resulting distribution is still cylindrical, because Gaussian distributions are additive and the bias information in the incoming message is irrelevant for the P2P message; the variance of the resulting distribution is the sum of measurement variance and the covariance of $\boldsymbol{p}_n^{(k)}$. Thus,

$$\begin{aligned}&\eta_{h_{n,m}\to \boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right) = \\ &p_{\mathcal{C}}\left(\boldsymbol{x}_m^{(k)}; r_{n\to m}^{(k)}, \boldsymbol{\mu}_{\boldsymbol{p}_{n\to m}^{(k)}}, \boldsymbol{\Sigma}_{\boldsymbol{p}_{n\to m}^{(k)}} + \sigma^2_{n\to m}\boldsymbol{I}\right),\end{aligned} \tag{6.26}$$

where $\boldsymbol{\mu}_{\boldsymbol{p}_{n\to m}^{(k)}}, \boldsymbol{\Sigma}_{\boldsymbol{p}_{n\to m}^{(k)}}$ are mean and variance of the message $\eta_{\boldsymbol{x}_n^{(k)} \to h_{n,m}}(\boldsymbol{x}_n^{(k)})$, $\boldsymbol{I}$ is the identity matrix, and

$$\begin{aligned}&p_{\mathcal{C}}\left(\boldsymbol{x}; \rho, \boldsymbol{\mu}_{\boldsymbol{p}_n}, \boldsymbol{\Sigma}_{\boldsymbol{p}_n+\rho}\right) \triangleq \\ &\tfrac{1}{Z}\exp\left[-\tfrac{1}{2}\left(\boldsymbol{p}' - \rho\tfrac{\boldsymbol{p}'}{\|\boldsymbol{p}'\|}\right)^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{p}_n+\rho}\left(\boldsymbol{p}' - \rho\tfrac{\boldsymbol{p}'}{\|\boldsymbol{p}'\|}\right)\right],\end{aligned} \tag{6.27}$$

where $\boldsymbol{p}' \triangleq \boldsymbol{p} - \boldsymbol{\mu}_{\boldsymbol{p}_n}$. A section of the *hyper-cylindrical* distribution family $\mathcal{C}$ is shown in Fig. 6.3c.

6.4.2.4 *Messages to peer-to-peer factors*

According to the SPA, messages from variable nodes to P2P factors are given by

$$\begin{aligned}\eta_{\boldsymbol{x}_m^{(k)}\to h_{n,m}}\left(\boldsymbol{x}_m^{(k)}\right) \propto\ &\eta_{f_m\to \boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right) \times \\ &\prod_{s\in\mathcal{S}_m^{(k)}} \eta_{g_{s,m}\to \boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right) \prod_{l\in\mathcal{M}_m^{(k)}\setminus n} \eta_{h_{l,m}\to \boldsymbol{x}_m^{(k)}}\left(\boldsymbol{x}_m^{(k)}\right),\end{aligned} \tag{6.28}$$

Due to heterogeneous shape and varying number of incoming messages, it is impossible to give a general expression of the resulting distribution. In addition, even if computed analytically, such distribution would be a very complex and irregular function, making it difficult to express messages at the

subsequent iteration. For these reasons, the result of message multiplication is approximated as a multivariate Gaussian distribution

$$\eta_{\boldsymbol{x}_m^{(k)} \to h_{n,m}}\left(\boldsymbol{x}_m^{(k)}\right) \approx p_{\mathcal{N}}\left(\boldsymbol{x}_m^{(k)}; \boldsymbol{\mu}_{\boldsymbol{x}_{m\to n}^{(k)}}, \boldsymbol{\Sigma}_{\boldsymbol{x}_{m\to n}^{(k)}}\right). \tag{6.29}$$

The problem then reverts to finding the parameters of the output distribution that best approximate the product of the incoming parametric messages. To this purpose, a parameter estimation algorithm is proposed (see Alg. 6.2) inspired from importance sampling [117]. At each iteration, samples $\boldsymbol{z}_i$ are drawn according to the current estimated output distribution, and weighted proportional to the ratio between the product of the probabilities of each of the incoming messages (computed according to their respective analytical distributions evaluated at the given samples) and the probability of the sample according to the output distribution. The distribution parameters are then refined through weighted sample mean and covariance estimators, and used to draw a new set of samples in the next iteration. No restrictions on the covariance matrix are imposed, so that it can take any ellipsoidal shape.

Algorithm 6.2: Parametric Message Multiplication

input : Initial estimate of $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$; distributions of all incoming messages (factors)

output: Updated $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$ after product

repeat

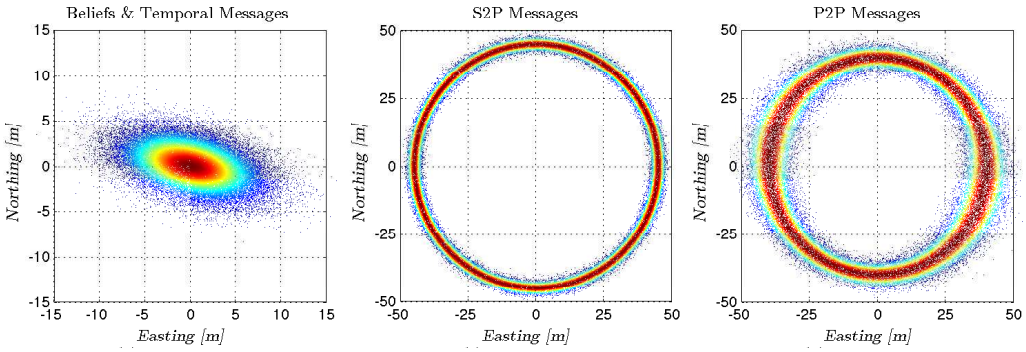
- Draw N samples $\boldsymbol{z}_i$ from $\mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\right)$
- Compute the probability of each sample $q\left(\boldsymbol{z}_i\right)$ in the distribution it was drawn from, using (6.21)
- Evaluate the PDFs at samples $p_j\left(\boldsymbol{z}_i\right)$ for each factor, using (6.21), (6.24), or (6.27)
- Assign a weight to each sample as: $w_i = \frac{\prod_j p_j(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)}$, then normalize them such that $\sum_{i=1}^{N} w_i = 1$
- Estimate new mean and covariance using weighted samples: $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}} = \sum_{i=1}^{N} w_i \boldsymbol{z}_i$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} = \frac{\sum_{i=1}^{N} w_i (\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\boldsymbol{x}})(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\boldsymbol{x}})^T}{1 - \sum_{i=1}^{N} w_i^2}$

until convergence

6.4.2.5 *Beliefs*

Finally, beliefs $\hat{p}(\boldsymbol{x}_m^{(k)})$, i.e. approximations of the marginals of interest $p(\boldsymbol{x}_m^{(k)} | \mathbb{p}_{\mathcal{M}}^{(1:k)}, \mathbb{r}_{\mathcal{M}}^{(1:k)})$ (6.13), are given by

$$\hat{p}(\boldsymbol{x}_m^{(k)}) \propto \eta_{f_m \to \boldsymbol{x}_m^{(k)}}(\boldsymbol{x}_m^{(k)}) \prod_{s \in \mathcal{S}_m^{(k)}} \eta_{g_{s,m} \to \boldsymbol{x}_m^{(k)}}(\boldsymbol{x}_m^{(k)}) \times \prod_{n \in \mathcal{M}_m^{(k)}} \eta_{h_{n,m} \to \boldsymbol{x}_m^{(k)}}(\boldsymbol{x}_m^{(k)}). \tag{6.30}$$



(a) $\mathcal{N}$ distribution. (b) $\mathcal{D}$ distribution. (c) $\mathcal{E}$ distribution.

Figure 6.4: Distribution families for the separate position bias model. Horizontal sections.

The above expression is a message multiplication similar to the one used for messages to P2P factors. Hence, the same parametric message multiplication algorithm (Alg. 6.2) can be used to compute beliefs:

$$\hat{p}\left(\boldsymbol{x}_m^{(k)}\right) \approx p_{\mathcal{N}}\left(\boldsymbol{x}_m^{(k)}; \boldsymbol{\mu}_{\boldsymbol{x}_m^{(k)}}, \boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k)}}\right). \tag{6.31}$$

6.4.3 *Parametric BP: Separate Position-Bias Model*

The message update rules for the FG of Fig. 6.2b are now described.

6.4.3.1 *Temporal messages*

Similar to the joint case, temporal messages are approximated by Gaussian distribution, but separately for bias and position variables:

$$\eta_{f_{\boldsymbol{p}_m} \to \boldsymbol{p}_m^{(k)}}\left(\boldsymbol{p}_m^{(k)}\right) \approx p_{\mathcal{N}}\left(\boldsymbol{p}_m^{(k)}; \boldsymbol{\mu}_{\boldsymbol{p}_m^{(k)}}, \boldsymbol{\Sigma}_{\boldsymbol{p}_m^{(k)}}\right), \tag{6.32}$$

$$\eta_{f_{b_m} \to b_m^{(k)}}\left(b_m^{(k)}\right) \approx p_{\mathcal{N}}\left(b_m^{(k)}; \mu_{b_m^{(k)}}, \sigma^2_{b_m^{(k)}}\right), \tag{6.33}$$

where $\boldsymbol{\mu}_{\boldsymbol{p}_m^{(k)}}$ and $\mu_{b_m^{(k)}}$ are computed, respectively, from mobility and clock drift models, and $\boldsymbol{\Sigma}_{\boldsymbol{p}_m^{(k)}}$ and $\sigma^2_{b_m^{(k)}}$ are updated such that $\boldsymbol{\Sigma}_{\boldsymbol{p}_m^{(k)}} \succeq \boldsymbol{\Sigma}_{\boldsymbol{p}_m^{(k-1)}}$, $\sigma^2_{b_m^{(k)}} \geqslant \sigma^2_{b_m^{(k-1)}}$ to take into account prediction uncertainty.

6.4.3.2 *Messages from satellite factors*

Messages from factors $g_{s,m}$ can now be divided into two types: messages to position variables, that take as input an estimated bias value, and messages to bias variables, which, conversely, take as input a position value. Based on

similar considerations as in Sec. 6.4.2.3, *satellite-to-position* messages can be computed as

$$\begin{aligned}&\eta_{g_{s,m}\to\boldsymbol{p}_m^{(k)}}\left(\boldsymbol{p}_m^{(k)}\right)=\\&p_{\mathcal{D}}\left(\boldsymbol{p}_m^{(k)};\rho_{s\to m}^{(k)}-\mu_{b_m^{(k)}},\boldsymbol{p}_s^{(k)},\sigma_{s\to m}^2+\sigma_{b_m^{(k)}}^2\right),\end{aligned}\tag{6.34}$$

where

$$p_{\mathcal{D}}\left(\boldsymbol{p};\rho,\boldsymbol{\mu}_s,\sigma_\rho^2\right)\triangleq\frac{1}{Z}\exp\left[-\frac{1}{2\sigma_\rho^2}\left(\|\boldsymbol{p}-\boldsymbol{\mu}_x\|-\rho\right)^2\right]\tag{6.35}$$

denotes a Gaussian distribution with *spherical* support (cf. [68]), characterized by radius ρ (pseudorange measurement), center $\boldsymbol{\mu}_s$ (satellite position), and variance σ_ρ^2. An example of $\mathcal{D}$ distribution is depicted in Fig. 6.4b.

Satellite-to-bias messages are represented by a univariate Gaussian PDF:

$$\begin{aligned}&\eta_{g_{s,m}\to b_m^{(k)}}\left(b_m^{(k)}\right)=\\&p_{\mathcal{N}}\left(b_m^{(k)};\rho_{s\to m}^{(k)}-\left\|\boldsymbol{p}_s^{(k)}-\boldsymbol{\mu}_{\boldsymbol{p}_m^{(k)}}\right\|,\sigma_{s\to m}^2+\mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{p}_m^{(k)}})\right),\end{aligned}\tag{6.36}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator.

6.4.3.3 *Messages from peer-to-peer factors*

Like it the joint case, P2P messages carry only position information. The cylindrical distribution therefore reduces to an *ellipsoidal* distribution in the $\boldsymbol{p}$ space, i.e., a section of the $\mathcal{C}$ distribution perpendicular to the b axis:

$$\begin{aligned}&\eta_{h_{n,m}\to\boldsymbol{p}_m^{(k)}}\left(\boldsymbol{p}_m^{(k)}\right)=\\&p_{\mathcal{E}}\left(\boldsymbol{p}_m^{(k)};r_{n\to m}^{(k)},\boldsymbol{\mu}_{\boldsymbol{p}_{n\to m}^{(k)}},\boldsymbol{\Sigma}_{\boldsymbol{p}_{n\to m}^{(k)}}+\sigma_{n\to m}^2\boldsymbol{I}\right),\end{aligned}\tag{6.37}$$

where

$$\begin{aligned}&p_{\mathcal{E}}(\boldsymbol{p};\rho,\boldsymbol{\mu}_{\boldsymbol{p}_n},\boldsymbol{\Sigma}_{\boldsymbol{p}_n+\rho})\triangleq\\&\frac{1}{Z}\exp\left[-\frac{1}{2}\left(\boldsymbol{p}'-\rho\frac{\boldsymbol{p}'}{\|\boldsymbol{p}'\|}\right)^T\boldsymbol{\Sigma}_{\boldsymbol{p}_n+\rho}^{-1}\left(\boldsymbol{p}'-\rho\frac{\boldsymbol{p}'}{\|\boldsymbol{p}'\|}\right)\right].\end{aligned}\tag{6.38}$$

with $\boldsymbol{p}'=\boldsymbol{p}-\boldsymbol{\mu}_{\boldsymbol{p}_n}$. The $\mathcal{E}$ distribution is depicted in Fig. 6.4c. Observe that, for $\rho=0$, it becomes a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{p}_n},\boldsymbol{\Sigma}_{\boldsymbol{p}_n+\rho})$.

6.4.3.4 *Messages to peer-to-peer factors*

These messages, having a product form as

$$\begin{aligned}\eta_{\boldsymbol{p}_m^{(k)}\to h_{m,n}}\left(\boldsymbol{p}_m^{(k)}\right)\propto\eta_{f_{\boldsymbol{p}_m}\to\boldsymbol{p}_m^{(k)}}\left(\boldsymbol{p}_m^{(k)}\right)\times\\\prod_{s\in\mathcal{S}_m^{(k)}}\eta_{g_{s,m}\to\boldsymbol{p}_m^{(k)}}\left(\boldsymbol{p}_m^{(k)}\right)\prod_{l\in\mathcal{M}_m^{(k)}\setminus n}\eta_{h_{l,m}\to\boldsymbol{p}_m^{(k)}}\left(\boldsymbol{p}_m^{(k)}\right),\end{aligned}\tag{6.39}$$

are approximated by D-variate Gaussian distributions $\mathcal{N}$, whose parameters are computed via the parametric message multiplication algorithm (Alg. 6.2).

6.4.3.5 *Messages to satellite factors*

These messages are unique to the separate representation. They include *bias-to-satellite messages*,

$$\eta_{b_m^{(k)} \to g_{s,m}}\left(b_m^{(k)}\right) \propto \eta_{f_{b_m} \to b_m^{(k)}}\left(b_m^{(k)}\right) \prod_{v \in S_m^{(k)} \setminus s} \eta_{g_{v,m} \to b_m^{(k)}}\left(b_m^{(k)}\right), \tag{6.40}$$

which can be computed as univariate Gaussian distributions with

$$\sigma^2_{b_m^{(k)} \to g_{s,m}} = \left[\frac{1}{\sigma^2_{f_{b_m} \to b_m^{(k)}}} + \sum_{y \in S_m^{(k)} \setminus s} \frac{1}{\sigma^2_{g_{y,m} \to b_m^{(k)}}} \right]^{-1}, \tag{6.41}$$

$$\mu_{b_m^{(k)} \to g_{s,m}} = \sigma^2_{b_m^{(k)} \to g_{s,m}} \left[\frac{\mu_{f_{b_m} \to b_m^{(k)}}}{\sigma^2_{f_{b_m} \to b_m^{(k)}}} + \sum_{y \in S_m^{(k)} \setminus s} \frac{\mu_{g_{y,m} \to b_m^{(k)}}}{\sigma^2_{g_{y,m} \to b_m^{(k)}}} \right], \tag{6.42}$$

and *position-to-satellite messages* $\eta_{p_m^{(k)} \to g_{s,m}}$, which can be approximated as Gaussian distributions using Alg. 6.2, similar to (6.39).

6.4.3.6 *Beliefs*

Position beliefs $\hat{p}(p_m^{(k)})$ are then computed like (6.39), including all incoming P2P messages $n \in \mathcal{M}_m$. *Bias* beliefs $\hat{p}(b_m^{(k)})$ are computed like (6.40), i.e., univariate Gaussian distributions with mean and variance given by (6.41,6.42), including all satellites $s \in \mathcal{S}_m$ in the summation.

6.4.4 *Complexity*

Table 6.1: Complexity of H-SPAWN for a single agent.

Computation	**No. of FLOP**
Compute time message	$2(D+1)^2 + 2(D+1)^3$
The following operations are repeated $P+1$ *times and iterated* I *times*	
Draw multiplication samples	$N(D+1)^2$
Evaluate message distributions	$N(S+P+2)(D+1)^3$
Compute message multiplication	$N(S+P+4)$
Estimate mean	$2N(D+1)$
Estimate covariance	$N(D+1)^2$

Complexity of the H-SPAWN algorithm is dominated by message multiplication needed in messages to P2P factors and messages from position variables to satellite factors. Considering a certain node agent with one temporal message, P peer-to-peer messages and S satellite-to-peer messages, using N

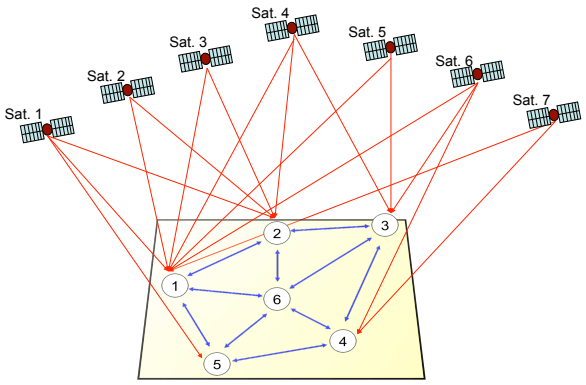


Figure 6.5: Example network topology.

samples z_i and requiring I iterations in the product estimate, the complexity scales as $\mathcal{O}(\mathrm{IPN}(S+P+1)(D+1)^3)$.In contrast, all other messages are computed analytically in $\mathcal{O}((S+P+1)(D+1))$. Table 6.1 lists the number of floating point operations per second (FLOP) needed in one execution of Alg. 6.2.

6.5 SIMULATION RESULTS

6.5.1 *Cramér-Rao Bound*

The analytical results derived in the Sec. 6.2 are now illustrated by a practical example. Consider the network depicted in Fig. 6.5, with six agents arranged in a star topology. Each agent can communicate with two neighbors, except agent 6, located in the center, that can communicate with all other agents. Agent 1 has visibility of all satellites; agent 2 can see four (the minimum number needed to estimate position and bias unambiguously); agents 3, 4, 5, and 6, on the contrary, are only connected to three, two, one, and no satellites, respectively. This configuration is representative of a network located in an indoor environment, where only agents close to windows or outer walls can receive satellite measurements.

Position of agent 6 (45.06° lat., 7.66° long., 311.96 m height) is taken as the origin of the reference system; the relative positions of the other agents, expressed in easting-northing-up (ENU) coordinates, are:

Agent no.	1	2	3	4	5	6
E [m]	-50	0	50	30	-30	0
N [m]	10	30	40	-20	-40	0
U [m]	0.27	0.92	-1.13	0.43	0.15	0

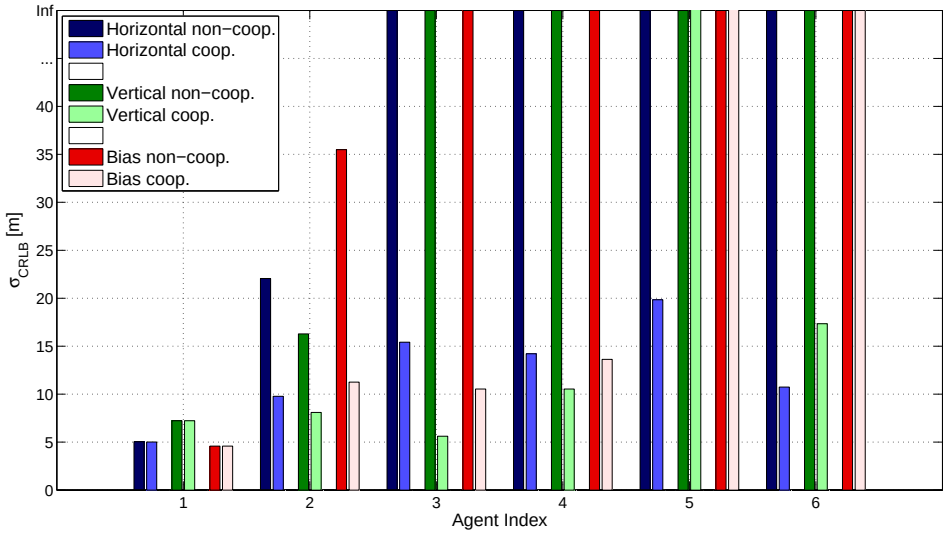


Figure 6.6: Comparison of position- and bias-CRB in realistic 3D scenario: non-cooperative (GNSS only) vs. hybrid (GNSS + peer-to-peer communication) setting.

Satellites' positions are drawn according to real GPS orbits. Their values, again expressed in ENU coordinates with respect to agent 6, are:

Sat. no.	1	2	3	4	5	6	7
E $[\cdot 10^6 m]$	-16.17	-9.18	-1.71	-13.97	14.28	22.95	-12.90
N $[\cdot 10^6 m]$	-4.02	-18.36	-10.50	10.83	6.46	4.86	21.68
U $[\cdot 10^6 m]$	14.02	10.78	18.15	13.31	15.01	5.83	2.44

The variance of pseudorange and range measurements is set, respectively, to $\sigma_{s\to m} = 5$ m $\forall m \in \mathcal{M}, s \in \mathcal{S}_m$ and $\sigma_{n\to m} = 0.20$ m $\forall m \in \mathcal{M}, n \in \mathcal{M}_m$. For simplicity no anchors are considered in this example.

Under this setting, the CRB is computed to compare the achievable positioning accuracy in the non-cooperative and in the hybrid scenario. Let $\boldsymbol{J}$ be the CRB matrix obtained by inversion of $\boldsymbol{F}_{\text{non-coop}}$ (6.10) or $\boldsymbol{F}$ (6.11), after removing rows and columns corresponding to non-estimable variables[4], and denote by $\boldsymbol{J}_m$ the $(D+1)\times(D+1) = 4\times 4$ block of $\boldsymbol{J}$ corresponding to agent m. Then, the positioning accuracy for each agent m can be decomposed into: a *horizontal* component, i.e, the trace of the E-N block of $\boldsymbol{J}_m$,

$$\sigma_{\text{CRB-hor}}(m) \doteq \sqrt{\boldsymbol{J}_m[1,1] + \boldsymbol{J}_m[2,2]},$$

a *vertical* component

$$\sigma_{\text{CRB-vert}}(m) \doteq \sqrt{\boldsymbol{J}_m[3,3]},$$

4 Non-estimable variables are: positions and biases, for agents whose total number of connections is less than $D+1$; biases, for agents connected to no satellites. These variables generate matrix singularities, hence CRB $\to \infty$.

and a *bias* component

$$\sigma_{\text{CRB}-\text{bias}}(m) \doteq \sqrt{\boldsymbol{J}_m[4,4]}.$$

The unit of all components is meters.

These performance metrics, plotted in Fig. 6.6, illustrate the benefits arising from cooperation. With the exception of agent 1, which has full visibility of all the available GPS satellites, the other agents obtain a significant performance improvement in the hybrid case. For agent 2, which sees four satellites, the CRB reduces by one half. Agents 3, 4, and 5 in the non-cooperative case have less measurements than unknowns, hence their CRB $\to \infty$; when peer-to-peer communication is introduced, the CRB takes relatively low values. Cooperation thus proves to be essential in GPS-challenged environments. Agent 6, finally, is able to estimate its position thanks to peer-to-peer information, but cannot estimate its bias in any case: at least one satellite connection is necessary, since range measurements do not carry any information about clock bias.

6.5.2 *H-SPAWN Performance*

In this section the performance of the H-SPAWN algorithm, introduced in Sec. 6.4, is analyzed in two scenarios: a small static network, as a proof of concept, and a large mobile network, as a realistic case. The performance of H-SPAWN is compared in both scenarios with two classical approaches: a hybrid-cooperative weighted least squares (HC-WLS) algorithm, inspired on the iterative descent method proposed in [158] and extended to hybrid GNSS-terrestrial ranging like in [44]; and the hybrid-cooperative extended Kalman filter (HC-EKF) algorithm presented in [15].

6.5.2.1 *Small-scale Static Scenario*

The first scenario, shown in Fig. 6.7a, is defined as follows:

- Satellite positions (Tab. 6.3) according to real GPS satellite orbits, seen from the equator.
- 6 nodes deployed on the Earth surface over an area of 100×100 m, as in Table 6.2.
- Uniform initial distributions in a circle of radius 1 km. (a *warm-start* condition, i.e., nodes have prior knowledge of the region where they are located when turned on).
- Static nodes. As a consequence, the following model is used when updating temporal messages: $\boldsymbol{x}_m^{(k)} = \boldsymbol{x}_m^{(k-1)}$; $\boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k)}} = \boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k-1)}} + \sigma^2 \boldsymbol{I}$, with $\sigma^2 = 4$.

- Pseudorange and P2P range measurements are corrupted by AWGN with standard deviation $\sigma_{s\to m} = 5$ m and $\sigma_{n\to m} = 0.2$ m respectively, typical values for GNSS in open sky conditions and UWB-based ranging in line-of-sight (LoS).
- New measurements are generated at each time slot, and algorithms run 5 iterations per slot.

The performance of H-SPAWN, HC-EKF, and HC-WLS in the above scenario is evaluated in terms of root mean squared error (RMSE), computed from the error vectors $\boldsymbol{\varepsilon}_m^{(k)} \triangleq \boldsymbol{\mu}_{p_m^{(k)}} - \boldsymbol{p}_m^{(k)}$ for all agents $m \in \mathcal{M}$ over 100 Monte Carlo simulations. The global localization RMSE is reported for each node in Fig. 6.7b, comparing the three considered algorithms after five time slots. H-SPAWN in its joint position-bias version outperforms all the other algorithms, while HC-WLS provides the worst accuracy overall. In order to evaluate the benefit of cooperation, the performance achieved by nodes is analyzed in relation to their connectivity conditions. Nodes 1 and 2 (which have enough satellites in view) achieve the least errors. More interesting are the cases of peers 3-5, which have 3, 2 and 1 satellite in view respectively and are therefore unable to localize themselves using GNSS only. For these nodes, localization is made possible thanks to cooperation, although with larger errors compared to nodes with complete satellite visibility. Finally, node 6, even without visible satellites, achieves good performance, thanks to collaboration with the other nodes in the network. RMSE values are close to the Cramér-Rao lower bound, derived for the same scenario in [101], thus confirming the effectiveness of the proposed algorithm implementation.

A more detailed comparison is shown in Figs. 6.7c-6.7d, where the CDFs of the global errors over all nodes in the network are plotted at the first and the fifth time slot. A coordinate transformation from ECEF to ENU is performed in order to analyze errors in horizontal and vertical components, as it is usual in GNSS performance tests. It can be noticed how H-SPAWN-joint, after a single time slot, achieves superior accuracy than H-SPAWN-separate after 5 slots, thus providing faster convergence. With more time slots, estimates are enhanced thanks to the availability of new measurements, and H-SPAWN-joint still obtains superior performance compared to other algorithms with the same number of measurements. Differences are more evident in the horizontal component than in the vertical one, where the HC-EKF has a performance close to that of H-SPAWN-joint.

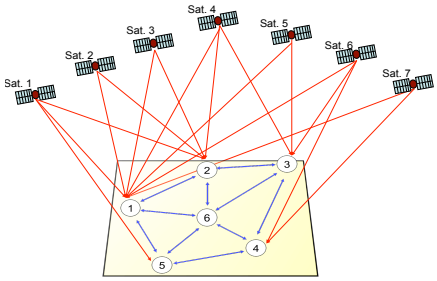
The convergence of the considered algorithms is shown in Figure 6.7e, which illustrates the evolution of the global RMSE iteration by iteration over 5 time slots. H-SPAWN-joint rapidly achieves a good performance in a few iterations, and slightly improves it with more measurements.

Table 6.2: Peer Positions in ENU for the small-scale simulation scenario.

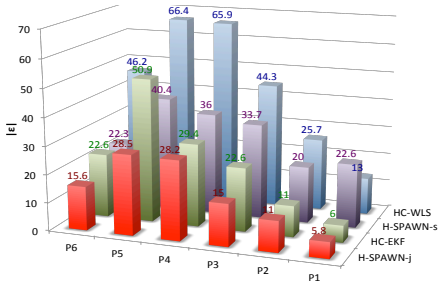
Peer	1	2	3	4	5	6
E[m]	-50	0	50	30	-30	0
N[m]	10	30	40	-20	-40	0
U[m]	-9	6	1	-7	2	0

Table 6.3: Satellite positions in ECEF coordinates.

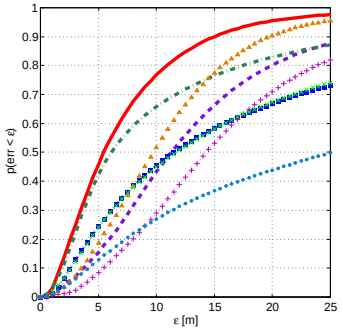
Sat.	1	2	3	4	5	6	7
x[m]	19263524	26124976	24768710	8048029	8543818	2082386	-7307117
y[m]	-13725770	-5749420	1601307	-13014437	15561017	23437415	-14002994
z[m]	11583188	-846377	9925575	21563572	19676844	12048145	21528812



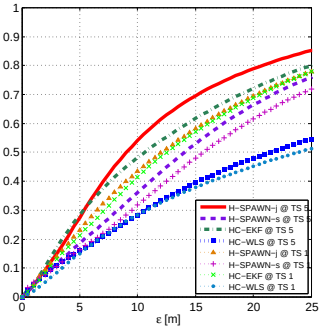
(a) Simulation scenario.



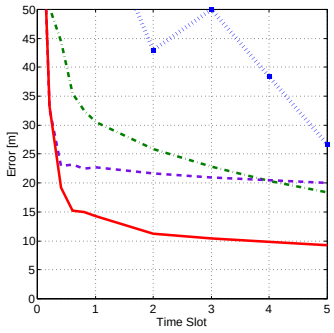
(b) Root mean squared errors.



(c) Horizontal errors CDF



(d) Vertical errors CDF



(e) Convergence of root mean square error.

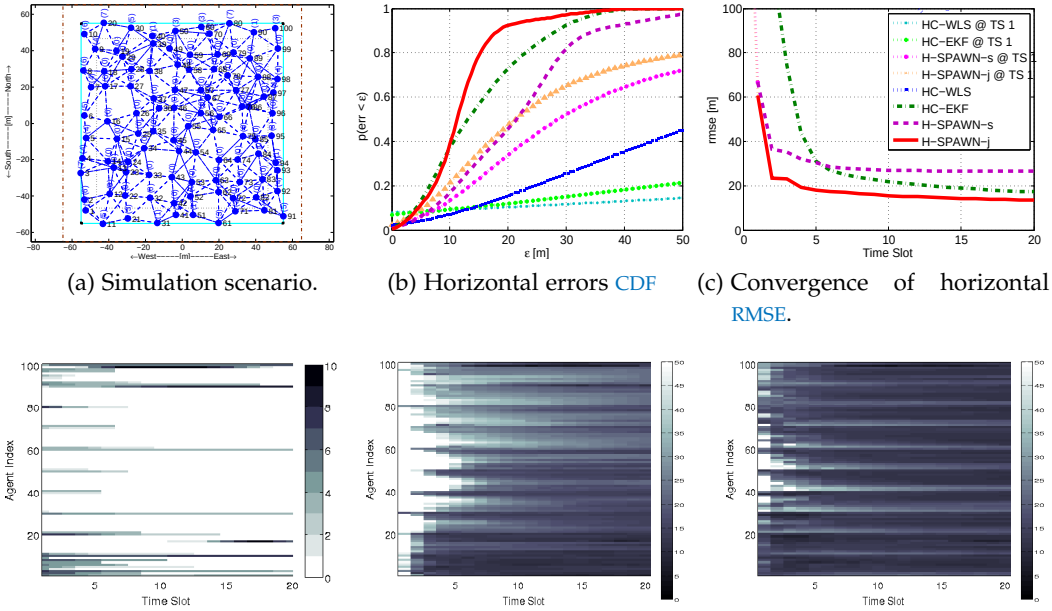
Figure 6.7: Small-scale simulation scenario and resulting RMSE by agent (top), global error CDF and RMSE evolution averaged over all agents (botom). Results after 1 and 5 time slots, 100 Monte Carlo runs.

6.5.2.2 *Large-scale Mobile Scenario*

In order to assess the performance of the proposed algorithm in a realistic, challenging scenario, a network of 100 mobile agents placed inside a building is now considered. Only nodes that are close to the walls have partial visibility of the GNSS constellation; in addition, it is assumed that nodes are able to communicate with peers if their distance is lower than a certain threshold, obtaining in this way a mesh network, with several hops between internal nodes (without any satellite visibility) and the outer ones (with partial visibility). The scenario is depicted in Fig. 6.8a and is defined as follows:

- Real GPS satellite orbits from 45° latitude.
- 100 nodes deployed on the Earth surface inside a building of 100×100 m, with *warm start* in a circle of radius 10 km.
- *Random-waypoint* mobility model [11]: each node moves towards a random point (selected with uniform probability) at constant velocity, uniformly chosen between 1 and 2m/s, remains in such point for a random time (uniform between 0 and 1s), and then moves to a new point. To make simulation more realistic, it is assumed that the mobility model is unknown to nodes. Therefore, temporal messages follow the usual model with $\boldsymbol{x}_m^{(k)} = \boldsymbol{x}_m^{(k-1)}$, $\boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k)}} = \boldsymbol{\Sigma}_{\boldsymbol{x}_m^{(k-1)}} + \sigma^2 \boldsymbol{I}$, with $\sigma^2 = 4$.
- Pseudorange are corrupted by AWGN with standard deviations varying from 10 to 20m depending on the satellite elevation angle seen from the peer (cf. [63]), modeling in this way *multipath fading* and *jitter* due to low carrier-to-noise ratio (C/N_0).
- P2P communication between nodes, say m and n, is possible only if $d_{m,n} \triangleq \|\boldsymbol{p}_m - \boldsymbol{p}_n\| < 20$m. Range measurements are corrupted by AWGN with $\sigma_{m\to n} = 0.2$m if $d_{m,n} \leqslant 10$m, or $\sigma_{m\to n} = 0.4$m if $10 < d_{m,n} \leqslant 20$m.
- New measurements are generated at each time slot. Algorithms are stopped after 3 iterations per slot. In order to better appreciate the effect of mobility, every Monte Carlo run consists now of 20slots with duration of 1s.

Fig. 6.8b shows the horizontal error c.d.f.s achieved by the considered algorithms in the above scenario, at the first and last time slot. Fig. 6.8c shows the evolution of the RMSE as a function of time slots. H-SPAWN in the joint version confirms its superiority compared to HC-WLS and HC-EKF. Remarkably, all algorithms exhibit a significant improvement of the estimated position accuracy with time, in spite of mobility (and consequent dynamic topology). Observe that H-SPAWN and HC-EKF achieve similar performance asymptotically in time (i.e., given a sufficient number of available measurements), but H-SPAWN provides faster convergence.



(a) Simulation scenario. (b) Horizontal errors CDF (c) Convergence of horizontal RMSE.

(d) Number of visible satellites by node. (e) Horizontal RMSE (m) by node using HC-EKF. (f) Horizontal RMSE (m) by node using H-SPAWN joint.

Figure 6.8: Large-scale mobile simulation scenario; numbers in brackets = number of visible satellites. Performance in terms of horizontal error CDF (top-center) and RMSE evolution (top-right) averaged over all agents, after 1 and 20 time slots. Bottom: node-by-node satellite visibility (left) and RMSE evolution using HC-EKF (center) or H-SPAWN (right).

Now, the performance achieved by individual nodes is analyzed. Fig. 6.8d shows the number of visible satellites for the nodes in the network. Satellite visibility slightly changes with time due to mobility, but in general it is denied for most nodes. Fig. 6.8e and Fig. 6.8f, then, illustrate the evolution of RMSE with time for each node. A strong correlation can be observed between these results and Fig. 6.8d: RMSE convergence is faster for nodes with good satellite visibility, whereas inner nodes need more time slots because they may be several hops away from GPS-enabled peers. However, for all nodes H-SPAWN provides shorter convergence times than HC-EKF. Therefore H-SPAWN proves more effective in exploiting peer-to-peer information exchange than competing algorithms.

7

MULTI-SENSOR SIGNAL DETECTION IN HETEROGENEOUS ENVIRONMENTS

7.1 MATHEMATICAL FORMULATION

7.1.1 *System Model*

Consider a cognitive network composed of K "secondary" users, coexisting with a "primary" network. The channel occupation for a certain secondary node k at time t is denoted by $\mathcal{H}_k^{(t)} \in \{0,1\}$. As explained in Sec. 1.1, a potentially heterogeneous detection setting is assumed, where the channel may be occupied by primary users in a certain location and free in another location. This situation may depend on the presence of physical barriers (e.g., walls), or on a predefined communication range above which interference is considered non-harmful (see Fig. 7.4).

The time is modeled as slotted. At each time slot, secondary users perform spectrum sensing by measuring the average energy, $\frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2$, where $\boldsymbol{y}_k^{(t)} \triangleq \left[y_k^{(t)}(1),\dots,y_k^{(t)}(N)\right]$ is a vector of N complex base-band received signal samples. Depending on $\mathcal{H}_k^{(t)}$, the generic sample can be written as

$$y_k^{(t)}(n) = \begin{cases} v_k(n) & \text{if} \quad \mathcal{H}_k^{(t)} = 0, \\ x_k(n) + v_k(n) & \text{if} \quad \mathcal{H}_k^{(t)} = 1, \end{cases} \tag{7.1}$$

where $v_k(n) \sim \mathcal{N}(0,\sigma_v^2)$ is complex white Gaussian noise, and $x_k(n)$ represents the signal received from a primary user if active. Signals are modeled as zero-mean random variables with $\mathbb{E}|x_k(n)|^2 \triangleq \sigma_k^2$, which includes the channel gain (assumed constant during the sensing period). The SNR at node k is defined as

$$\rho_k \triangleq \frac{\sigma_k^2}{\sigma_v^2}. \tag{7.2}$$

7.1.2 *Single Node Detection*

Since hypotheses $\mathcal{H}_k^{(t)}$ may be different for different nodes and at different time slots, the most general approach is obtained assuming all sensors as independent and performing signal detection separately at each single node and time slot. In this perspective, the test statistic available at node k is the vector $\boldsymbol{y}_k^{(t)}$. Assuming the signal samples as complex Gaussian distributed (which is a reasonable approximation also for PSK or QAM modulated signals,

taking into account pulse shape filter and non-coherent reception), the single-sensor likelihood ratio test (LRT) is given by

$$L_k^{(t)} \triangleq \frac{p(\boldsymbol{y}_k^{(t)}|\mathcal{H}_k^{(t)}=1)}{p(\boldsymbol{y}_k^{(t)}|\mathcal{H}_k^{(t)}=0)} \tag{7.3}$$

$$= \frac{(\sigma_v^2)^{N-1}}{1+\rho_k} \exp\left(\frac{\|\boldsymbol{y}_k^{(t)}\|^2}{\sigma_v^2}\frac{\rho_k}{1+\rho_k}\right) \underset{\hat{\mathcal{H}}_k^{(t)}=0}{\overset{\hat{\mathcal{H}}_k^{(t)}=1}{\gtrless}} \tau. \tag{7.4}$$

The above LRT can be seen from two points of view: as a Bayesian test, if the prior distributions of $\mathcal{H}_k^{(t)}$ and of $\boldsymbol{y}_k^{(t)}$ are uniform (non-informative); or as a Neyman-Pearson (NP) test, in which case the threshold τ is chosen as a function of the desired false alarm rate α. From (7.4) it follows that the energy $\frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2$ is a sufficient statistic for the LRT, i.e., energy detection is optimal in case of single-sensor detection. Thus, applying logarithms on both sides of L_k, the test can be rewritten as

$$\frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2 \underset{\hat{\mathcal{H}}_k^{(t)}=0}{\overset{\hat{\mathcal{H}}_k^{(t)}=1}{\gtrless}} \underbrace{\frac{\sigma_v^2}{N}\left(1+\frac{1}{\rho_k}\right)\left[\log(1+\rho_k)+\log\tau-(N-1)\log\sigma_v^2\right]}_{\triangleq\eta}. \tag{7.5}$$

If the NP approach is adopted, the new threshold η needs not be computed from τ, but it is chosen such that $\Pr\left(\frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2 > \eta \middle| \mathcal{H}_k^{(t)}=0\right) = \alpha$, which yields

$$\eta(\alpha) = \sigma_v^2\left[1+N^{-1/2}Q^{-1}(\alpha)\right], \tag{7.6}$$

as follows from well-known results on energy detection (cf. [67, 80]).

Then, define for each node an *individual observation factor*

$$F_k^{(t)} \triangleq \frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2 - \eta(\alpha) \underset{\hat{\mathcal{H}}_k^{(t)}=0}{\overset{\hat{\mathcal{H}}_k^{(t)}=1}{\gtrless}} 0. \tag{7.7}$$

Note that the above statistics have the form of log-likelihood ratio (LLR) tests. Also, thanks to the adoption of a NP approach, factors F_k only require knowledge of the noise variance (which can be estimated as the average energy of no-signal slots, offline or online) but not of the SNR of the signal to be detected. On the contrary, Bayesian likelihood ratios (e.g., [66, 109]) require prior knowledge of the SNR under hypothesis $\mathcal{H}=1$ and lack a general analytic method to set the decision threshold as a function of the false-alarm rate.

7.1.3 *Statistical Dependencies*

Single-user detection is optimal only when all sensors are uncorrelated. In realistic scenarios, it is likely that some correlation exists between the channel

occupancy of neighboring nodes, although such correlation is *a priori* unknown and may be time-varying (e.g., due to mobile primary users, changes in the radio environment, etc.). This condition can be taken into account through a *pairwise Markov random field (MRF)* model. As such, a joint prior distribution of variables $\mathcal{H}_k^{(t)}$ is introduced in the form of a product of pairwise exponential terms:

$$p(\mathcal{H}_1^{(t)}, \ldots, \mathcal{H}_K^{(t)}) = \frac{1}{Z} \prod_{\substack{j \in \mathcal{N}_K \\ k<j}} \exp\left(\lambda \Delta_{kj}^{(t)} \cdot \mathbf{1}\{\mathcal{H}_k^{(t)} = \mathcal{H}_j^{(t)}\}\right), \tag{7.8}$$

where $\mathcal{N}_k$ is the set of nodes within the communication range of k ("neighbors"), Z is a normalization constant such that the probability sums to 1, $\lambda < 1$ is a small positive constant, $\mathbf{1}\{x\} = 1$ if x is true or 0 otherwise, and $\Delta_{kj}^{(t)}$ is learned from a number T of previous time slots as follows:

$$\Delta_{kj}^{(t)} \triangleq \sum_{q=t-T}^{t-1} \mathbf{1}\{\hat{\mathcal{H}}_k^{(q)} = \hat{\mathcal{H}}_j^{(q)}\} - \mathbf{1}\{\hat{\mathcal{H}}_k^{(q)} \neq \hat{\mathcal{H}}_j^{(q)}\}. \tag{7.9}$$

In practice, $\Delta_{kj}^{(t)}$ can be updated recursively by

$$\Delta_{kj}^{(t)} = \Delta_{kj}^{(t-1)} + \mathbf{1}\{\hat{\mathcal{H}}_k^{(t-1)} = \hat{\mathcal{H}}_j^{(t-1)}\} - \mathbf{1}\{\hat{\mathcal{H}}_k^{(t-1)} \neq \hat{\mathcal{H}}_j^{(t-1)}\}. \tag{7.10}$$

The rationale for the proposed model is the following: the exponential terms in (7.8) assign higher probability to the event of nodes k and j having the same state $\mathcal{H}_k^{(t)} = \mathcal{H}_j^{(t)}$, if the decisions of nodes k and j have been equal in the majority of previous observations. The strength of interconnection between $\mathcal{H}_k^{(t)}$ and $\mathcal{H}_j^{(t)}$ is adjusted by the product $\lambda\Delta_{kj}$. Since $-T < \Delta_{kj} < T$, the constant λ should be chosen such that $\exp(\lambda T) \gg 1$, i.e., a large number of equal decisions between nodes k and j results in a high probability of $\mathcal{H}_k^{(t)}$ and $\mathcal{H}_j^{(t)}$ being equal. On the other hand, if $|\Delta_{kj}|$ is small (no significant correlation between previous decisions of nodes k and j), then $\exp(\lambda\Delta_{kj}) \approx 1$, i.e., the MRF distribution becomes non-informative.

7.1.4 *Resulting Model*

By combining the prior MRF joint distribution with individual observation likelihoods, the joint *a posteriori* distribution of variables $\mathcal{H}_1^{(t)}, \ldots, \mathcal{H}_K^{(t)}$ can be expressed as

$$p(\mathcal{H}_1^{(t)}, \ldots, \mathcal{H}_K^{(t)} | \boldsymbol{y}_1^{(t)}, \ldots, \boldsymbol{y}_K^{(t)}) \propto \prod_{k=1}^{K} p(\boldsymbol{y}_k^{(t)} | \mathcal{H}_k^{(t)}) \prod_{\substack{j \in \mathcal{N}_k \\ k<j}} \exp\left(\lambda \Delta_{kj}^{(t)} \cdot \mathbf{1}\{\mathcal{H}_k^{(t)} = \mathcal{H}_j^{(t)}\}\right). \tag{7.11}$$

Note that, if $p(\mathcal{H}_k^{(t)})$ is uniform, then $p(\boldsymbol{y}_k^{(t)}|\mathcal{H}_k^{(t)})$ and $p(\mathcal{H}_k^{(t)}|\boldsymbol{y}_k^{(t)})$ are equivalent up to a proportionality constant. Then, the signal detection problem for every node k can be formulated as a LRT of the marginal a posteriori probabilities,

$$L_k^{\text{post}} \triangleq \frac{p(\mathcal{H}_k^{(t)} = 1|\boldsymbol{y}_1^{(t)}, \dots, \boldsymbol{y}_K^{(t)})}{p(\mathcal{H}_k^{(t)} = 0|\boldsymbol{y}_1^{(t)}, \dots, \boldsymbol{y}_K^{(t)})}, \tag{7.12}$$

with

$$p(\mathcal{H}_k^{(t)}|\boldsymbol{y}_1^{(t)}, \dots, \boldsymbol{y}_K^{(t)}) = \sum_{\sim\{\mathcal{H}_k^{(t)}\}} p(\mathcal{H}_1^{(t)}, \dots, \mathcal{H}_K^{(t)}|\boldsymbol{y}_1^{(t)}, \dots, \boldsymbol{y}_K^{(t)}) \tag{7.13}$$

where the notation $\sum_{\sim\{\mathcal{H}_k^{(t)}\}}$ denotes multiple summation over all variables $\{\mathcal{H}_1^{(t)}, \dots, \mathcal{H}_K^{(t)}\}$ except $\mathcal{H}_k^{(t)}$.

The next section introduces a decentralized algorithm which, through an iterative exchange of local messages between network nodes, approximates the marginal LRTs of interest (7.12) while at the same time setting constraints on the local false-alarm probabilities for each node, in a NP-like fashion.

7.2 MESSAGE PASSING ALGORITHM

In order to estimate the marginal LRTs, the sum-product algorithm over factor graphs [62] is applied to the model defined by Eq. (7.11). In the considered model, each factor is connected to at most two variables, therefore message update rules can be expressed directly from one variable to another. Note that a decentralized implementation of the BP algorithm is enabled by the one-to-one correspondence between variables $\mathcal{H}_k^{(t)}$ and nodes in the network. As such, a generic message from node k to node j at iteration l is given by

$$\mu_{k\to j}^{(t,l)}(\mathcal{H}_j^{(t)}) \propto \sum_{\mathcal{H}_k^{(t)}\in\{0,1\}} \left(p(\boldsymbol{y}_k^{(t)}|\mathcal{H}_k^{(t)}) \exp(\lambda \Delta_{kj}^{(t)} \cdot \mathbf{1}\{\mathcal{H}_k^{(t)} = \mathcal{H}_j^{(t)}\}) \prod_{n\in\mathcal{N}_k\setminus\{j\}} \mu_{n\to k}^{(t,l-1)}(\mathcal{H}_k^{(t)}) \right). \tag{7.14}$$

Messages at iteration $l = 0$ are initialized as uniform distributions. Beliefs, i.e., estimates of the marginal probabilities (7.13), are computed at each iteration l as

$$b_k^{(t,l)}(\mathcal{H}_k^{(t)}) \propto \prod_{n\in\mathcal{N}_k} \mu_{n\to k}^{(t,l)}(\mathcal{H}_k^{(t)}). \tag{7.15}$$

The proportionality sign in (7.14) and (7.15) indicates that beliefs and messages are expressed up to a constant, which can be found by normalizing $b(\cdot)$ and $\mu(\cdot)$ so as to sum to 1.

Message update rules are now expressed in terms of LLRs. Define LLR beliefs

$$\Lambda_k^{(t,l)} \triangleq \log \frac{b_k^{(t,l)}(\mathcal{H}_k^{(t)}=1)}{b_k^{(t,l)}(\mathcal{H}_k^{(t)}=0)} \tag{7.16}$$

and messages

$$M_{k\to j}^{(t,l)} \triangleq \log \frac{\mu_{k\to j}^{(t,l)}(\mathcal{H}_j^{(t)}=1)}{\mu_{k\to j}^{(t,l)}(\mathcal{H}_j^{(t)}=0)}. \tag{7.17}$$

After algebraic manipulations, Message and belief update rules can be rewritten in LLR form as

$$M_{k\to j}^{(t,l)} = \mathcal{S}\left(\lambda\Delta_{kj}^{(t)},\ \log L_k^{(t)} + \sum_{n\in\mathcal{N}_k\setminus\{j\}} M_{n\to k}^{(t,l-1)}\right), \tag{7.18}$$

$$\Lambda_k^{(t,l)} = \log L_k^{(t)} + \sum_{n\in\mathcal{N}_k} M_{n\to k}^{(t,l)}, \tag{7.19}$$

where

$$\mathcal{S}(a,b) \triangleq \log \frac{1+e^{a+b}}{e^a+e^b}, \tag{7.20}$$

and $L_k^{(t)}$ is given by (7.3). As explained in Sec. 7.1.2, expressing $L_k^{(t)}$ in explicit form requires knowledge of the distribution of $\boldsymbol{y}_k^{(t)}$ under the hypothesis of signal present, which ultimately amounts to the knowledge of the SNR ρ_k. Furthermore, even assuming a prior knowledge, or guess, of the signal strength, the problem of setting the decision threshold τ remains unsolved: if one assigns equal weights to type-I and type-II errors (false alarms and missed detections), then the threshold should be set to 1 (i.e., 0 in the LLR domain), but this approach is not useful to ensure a false-alarm rate lower than a specified value.

On the other hand, it has been3 shown in Sec. 7.1.2 that the LRT on $L_k^{(t)}$ is equivalent to a LLR test of the energy $\frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2$ against a modified threshold $\eta(\alpha)$ that can be expressed directly as a function of the false-alarm rate. Therefore, set

$$\log L_k^{(t)} := F_k^{(t)} = \frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2 - \eta(\alpha) \tag{7.21}$$

in (7.18) and (7.19), with $F_k^{(t)}$ defined in (7.7). In this way a combined NP-Bayesian methodology has been applied to the multi-sensor detection problem: individual observations are processed by a NP approach, which is insensitive to SNR knowledge and offers control over the false-alarm rate, and are reinforced by peer-to-peer collaboration, implemented by means of a Bayesian joint prior distribution estimated from previous time slots.

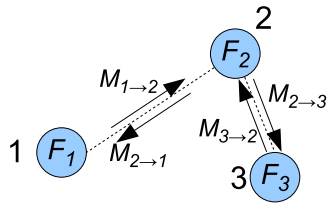


Figure 7.1: Example of a factor graph for a simple network of 3 nodes, representing messages and individual likelihood functions at a certain time slot t and iteration l. The statistical graph matches with the network topology, therefore messages can be exchanged by nodes in a decentralized fashion.

After a sufficient number of iterations (L), each node makes a decision on $\mathcal{H}_k^{(t)}$ simply based on the sign of its LLR belief:

$$\Lambda_k^{(t,L)} \underset{\hat{\mathcal{H}}_k^{(t)}=0}{\overset{\hat{\mathcal{H}}_k^{(t)}=1}{\gtrless}} 0. \tag{7.22}$$

Observe that the threshold on beliefs is set to zero as a result of (7.7). This implies that for each node the achieved false-alarm probability is always equal to α (in case of no correlation with neighbors) or lower. More details about this property are given in Sec. 7.3.

The graphical model is illustrated in Fig. 7.1 for a simple network of $K = 3$ nodes. Since all factors are pairwise, factors are simply represented as edges in the graph, and variables as vertices. This representation allows an intuitive mapping between graphical model and physical network. The figure shows the intrinsic information of each node, F_k, which can interpreted as a factor connected to a single variable, and messages $M_{k\to j}$ exchanged in the network at a certain iteration and time slot.

The resulting algorithm, called *NP-BP*, is summarized in Alg. 7.1. From the point of view of implementation it is worth noting that messages defined in LLR form are scalar numbers, which can be represented using few bits in the packets that are exchanged in the network. Finally, the number of iterations needed to reach convergence of the algorithm depends, in general, on the size and on the structure of the "clusters" of the network, i.e., the groups of nodes under the same hypothesis. It is observed empirically that very few iterations (e.g., 3) are enough to reach convergence (at least of final decisions, i.e., binary beliefs) even in large-scale networks.

7.3 PERFORMANCE ANALYSIS

In this section the performance of the proposed NP-BP method is investigated. First of all observe that, for sufficiently large T, the quantity $\Delta_{kj}^{(t)}$ defined in (7.9) is directly related to the correlation $r_{kj}^{(t)}$ between variables $\mathcal{H}_k$ and $\mathcal{H}_j$[1]:

$$r_{kj}^{(t)} > 0 \quad \Leftrightarrow \quad \Delta_{kj}^{(t)} \to \infty \tag{7.23}$$

$$r_{kj}^{(t)} = 0 \quad \Leftrightarrow \quad \Delta_{kj}^{(t)} \to 0 \tag{7.24}$$

$$r_{kj}^{(t)} < 0 \quad \Leftrightarrow \quad \Delta_{kj}^{(t)} \to -\infty. \tag{7.25}$$

The first case includes in particular $\rho_{kj}^{(t)} = 1$, which means $\mathcal{H}_k = \mathcal{H}_j$ (two nodes under the same hypothesis). The second case represents nodes that observe uncorrelated data, for example activity from different primary users. The third case occurs if two nodes remain under opposite hypotheses for long time.

Then, it is useful to analyze the asymptotic behavior of the function $\mathcal{S}(a, b)$ defined by (7.20). Note that: *(i)* $\mathcal{S}(a, b)$ is symmetric with respects to its arguments, a and b; *(ii)* when one of the arguments is significantly larger than the other one, the function tends to $\min\{a, b\}$. The function $\mathcal{S}(a, b)$ is plotted in Fig. 7.2 for different values of the parameters. In the message update rule (7.18), the first argument is $\lambda\Delta_{kj}^{(t)}$, i.e., a quantity that grows with the number of available previous observations when these reveal correlation with neighboring nodes; the second argument is the sum of the neighbor's individual likelihood ratio and incoming messages.

Based on these preliminary considerations, two extreme operating conditions of the NP-BP can be identified.

1. *No available past observations* or *no correlation between nodes* k *and* j: $\lambda\Delta_{kj}^{(t)} \ll F_k^{(t)}$. In this case, from (7.18),

$$M_{k\to j}^{(t,l)} \approx \lambda\Delta_{kj}^{(t)}. \tag{7.26}$$

Then, assuming that a node k is uncorrelated from its neighbors $n \in \mathcal{N}_k$, it follows from (7.19) that

$$\Lambda_k^{(t,L)} \approx F_k^{(t)} + \lambda \sum_{n\in\mathcal{N}_k} \Delta_{nk}^{(t)}. \tag{7.27}$$

Since λ is a small constant (typically $\ll 1$) and, in absence of correlation between nodes, $\Delta_{nk}^{(t)} \approx 0$ (7.24), Eq. (7.27) reduces to $\Lambda_k^{(t,l)} \approx F_k^{(t)}$. Note that this approximation is motivated not only for long observation periods (large T), but also for small T, because by definition $\lambda\Delta_{nk}^{(t)}$ takes

1 The correlation coefficient is defined as $r_{kj}^{(t)} \triangleq \frac{\mathbb{E}[(\mathcal{H}_k-\mu_k)(\mathcal{H}_j-\mu_j)]}{\sigma_k\sigma_j} = \frac{\mathbb{E}[\mathcal{H}_k\mathcal{H}_j]-0.25}{0.25}$. The index t is dropped as $\mathcal{H}_k$ and $\mathcal{H}_j$ are considered here as random variables, of which $\mathcal{H}_k^{(t)}$ and $\mathcal{H}_j^{(t)}$ are realizations at time t.

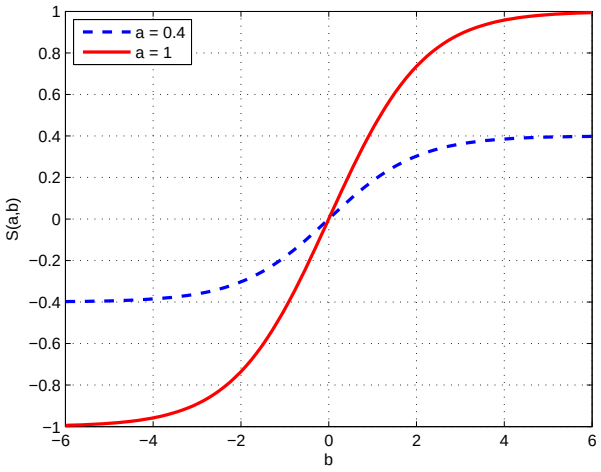


Figure 7.2: Plot of $S(a, b)$ for different values of a.

significant values only after a certain number of time slots. In addition, if there are enough neighbors having no correlation with k, the sign of $\Delta_{nk}^{(t)}$ is positive or negative with equal probability, hence the sum tends to zero.

As a result, in absence of correlation, the proposed spectrum sensing procedure *reduces to single-user energy detection*: each node separately tests $F_k^{(t)} \gtrless 0$, which is equal to $\frac{1}{N}\|\boldsymbol{y}_k^{(t)}\|^2 \gtrless \eta(\alpha)$. This property guarantees that the average false-alarm rate achieved by each node is, in the worst case (no cooperation among nodes), equal to α, thus satisfying the Neyman-Pearson requirement of an upper bound to the false-alarm rate.

In case of a sudden change in the radio environment that makes two previously correlated nodes (k, j) become uncorrelated, the system automatically learns the new situation within a few time slots thanks to the update of $\Delta_{kj}^{(t)}$ (7.9). The learning rate is determined by the coefficient λ: lower λ results in a slower adaptation and more conservative use of neighbors' data; higher λ makes the system more reactive to changes, but may result in overconfident estimation of correlations with neighbors. The optimization of λ is a subject beyond the scope of this paper. Empirically, the value of $\lambda = 0.1$ was chosen and used in simulations (see Sec. 7.4).

2. *Significant correlation between nodes* k *and* j: $\lambda\Delta_{kj}^{(t)} \gg F_k^{(t)}$. In this case, (7.18) becomes

$$M_{k\to j}^{(t,l)} \approx F_k^{(t)} + \sum_{n\in\mathcal{N}_k\setminus\{j\}} M_{n\to k}^{(t,l-1)}. \tag{7.28}$$

Applied iteratively, the above expression leads to formation of *clusters* of nodes correlated with each other: after a number of iterations L sufficient to span the entire cluster, the test statistic of each node in the cluster converges to the sum of individual LLR functions of all such nodes. More formally, let $\mathcal{C}$ be a set of nodes under the same hypothesis $\mathcal{H}_k^{(t)} = 0$ or $\mathcal{H}_k^{(t)} = 1$ $\forall k \in \mathcal{C}$ and $\forall t$ in a certain observation period of T slots; then, for any $k \in \mathcal{C}$,

$$\Lambda_k^{(t,L)} \approx \sum_{k\in\mathcal{C}} F_k^{(t)} = |\mathcal{C}| \left(\frac{1}{|\mathcal{C}|N} \sum_{k\in\mathcal{C}} \|\boldsymbol{y}_k^{(t)}\|^2 - \eta(\alpha) \right), \tag{7.29}$$

where $|\mathcal{C}|$ is the cardinality of $\mathcal{C}$, i.e., the number of nodes in the cluster. Hence, the test for any $k \in \mathcal{C}$ can be rewritten as

$$\frac{1}{|\mathcal{C}|N} \sum_{k\in\mathcal{C}} \|\boldsymbol{y}_k^{(t)}\|^2 \underset{\hat{\mathcal{H}}_k^{(t)}=0}{\overset{\hat{\mathcal{H}}_k^{(t)}=1}{\gtrless}} \eta(\alpha) \tag{7.30}$$

and is equivalent to *cooperative energy detection* by all nodes in the cluster. The advantage of the proposed method is that cooperation is achieved in a fully decentralized way, without a fusion center like in traditional cooperative spectrum sensing approaches. In addition, this method does not require prior knowledge of the nodes participating in cooperative detection, because clustering is determined automatically based on the evolution of coefficients $\Delta_{kj}^{(t)}$. These properties ensure robustness and scalability of the sensing algorithm.

7.3.1 *False-alarm Probability and Selection of the Threshold*

7.3.1.1 *Single-user threshold*

The threshold $\eta(\alpha)$ given by (7.6) is tailored for the single-user case. Clearly, if a node has no collaborating neighbors, the resulting false-alarm probability is $P_{fa} = \alpha$. However, cooperation among multiple nodes provides an improvement of the false-alarm probability at a given threshold: if all nodes in $\mathcal{C}$ are under the same condition ($\mathcal{H} = 0$), the resulting false-alarm probability is

$$P_{fa}(\mathcal{C}) = Q\left[\sqrt{|\mathcal{C}|N} \left(\frac{\eta}{\sigma_v^2} - 1 \right) \right], \tag{7.31}$$

as in an energy detector with $|\mathcal{C}|N$ samples. Therefore, setting η as (7.6) gives in the cooperative case a lower false-alarm rate than the nominal value α.

7.3.1.2 *Cooperative threshold*

In case of cooperation, a new threshold

$$\eta_{\mathcal{C}}(\alpha) = \sigma_v^2 \left[1 + (|\mathcal{C}|N)^{-1/2} Q^{-1}(\alpha) \right] \tag{7.32}$$

can be then computed from inversion of (7.31), resulting exactly in $P_{fa}(\mathcal{C}) = \alpha$ for all $k \in \mathcal{C}$. The new "cooperative" threshold can be set once clusters are established with sufficient stability, and the number of nodes in the cluster ($|\mathcal{C}|$) must be constantly verified based on coefficients $\Delta_{kj}^{(t)}$. Note that such a threshold selection may fail to guarantee the required false-alarm rate in case of changes in the radio environment for the nodes in $\mathcal{C}$. However, thanks to the automatic update of coefficient $\Delta_{kj}^{(t)}$, normal conditions are restored within a few time slots.

7.3.2 *Detection Performance*

The resulting detection probability for a cluster $\mathcal{C}$ can be computed again by using results from energy detection theory (cf. [67, 80]). From (7.30), the test statistic can be written as

$$T_{\mathcal{C}} \triangleq \frac{1}{|\mathcal{C}|N} \sum_{k\in\mathcal{C}} \|\boldsymbol{y}_k^{(t)}\|^2 = \frac{1}{|\mathcal{C}|} \sum_{k\in\mathcal{C}} T_k, \tag{7.33}$$

and $T_k \sim \mathcal{N}\left(1+\overline{\rho}_{\mathcal{C}}, \frac{1}{|\mathcal{C}|^2 N} \sum_{k\in\mathcal{C}} (1+\rho_k)^2\right)$, with $\overline{\rho}_{\mathcal{C}} \triangleq \frac{1}{|\mathcal{C}|} \sum_{k\in\mathcal{C}} \rho_k$, by linearity of the normal distribution. Hence, the detection probability is

$$P_d(\mathcal{C}) = \Pr[T_{\mathcal{C}} > \eta] = Q\left(|\mathcal{C}|\sqrt{N} \frac{\eta/\sigma_v^2 - 1 - \overline{\rho}_{\mathcal{C}}}{\sqrt{\sum_{k\in\mathcal{C}} (1+\rho_k)^2}} \right). \tag{7.34}$$

If all nodes in $\mathcal{C}$ are under the same SNR ρ, the above formula reduces to

$$P_d(\mathcal{C}) = Q\left[\sqrt{|\mathcal{C}|N} \left(\frac{\eta/\sigma_v^2}{1+\rho} - 1 \right) \right], \tag{7.35}$$

which is the same performance of a single energy detector with $|\mathcal{C}|N$ samples, similarly as in (7.31).

Note that, if a node has no cooperating neighbors, the same result (7.35) holds with $|\mathcal{C}| = 1$ thus reducing to traditional single-user energy detection.

7.3.3 *Complexity*

The computational complexity of NP-BP is very low, since in the adopted LLR representation all messages are scalar quantities, and their computation only involves sums and products. From a networking point of view, the main issue is message exchange, because *(i)* messages occupy some bandwidth (although limited), and *(ii)* packet transmission may fail and require retransmission(s). The latter problem is particularly relevant if nodes communicate over the same wireless channel that they are sensing, and that is possibly occupied by a primary user. In order to minimize both bandwidth usage and

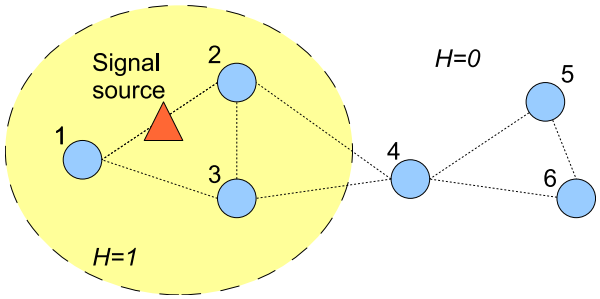


Figure 7.3: Small-scale network ($K = 6$ nodes). 3 nodes are under $\mathcal{H} = 1$, 3 under $\mathcal{H} = 0$. Each node can communicate with its neighbors; connections are indicated by dotted lines.

transmission errors, it is convenient to reduce the number of iterations of NP-BP. As an example, simulation results shown in the next section demonstrate that 3 iterations per time slot already provide excellent performance in all considered scenarios.

7.4 Simulation Results

The performance of the proposed algorithm is evaluated under two different simulation scenarios: the small network of Fig. 7.3 and the large, random network of Fig. 7.4.

The first case consists on $K = 6$ nodes; 3 of them experience the presence of a primary signal, while the other 3 nodes observe only noise. Each node can communicate with a limited number of neighbors, indicated by dotted lines in the figure. The second case is a large-scale network modeled as a "random geometric graph" [115], where nodes ($K = 50$) are deployed randomly in a square of size 100×100 units, and can communicate with each other if their distance is less than $r = 20$ units. It is assumed that nodes in the left part of the plane are under hypothesis $\mathcal{H} = 0$, while nodes on the right are in the presence of a signal.

Results for the small-scale scenario are reported in Figures 7.5, 7.6, and 7.7. The plots illustrate the false-alarm probability, computed for one of the nodes under $\mathcal{H} = 0$, namely node 5, and the detection probability, computed for one of the nodes under $\mathcal{H} = 1$, namely node 2, at three different time slots: $t = 1$, $t = 5$, and $t = 10$. For all slots, BP is stopped at the third iteration. Observe that at $t = 1$, the NP-BP algorithm reduces to single-user detection, because no observations are available, hence all variables $\Delta_{kj}^{(t)}$ are equal to 0.

In Fig. 7.5, the SNR is constant (-5 dB) for all nodes under hypothesis $\mathcal{H} = 1$. The graph shows that NP-BP significantly improves the performance of single-user energy detection (i.e., results at $t = 1$), both in terms of false-alarm and of detection probabilities. The gap increases with time, as the system gradually learns the correlation parameters. At $t = 10$, the detection

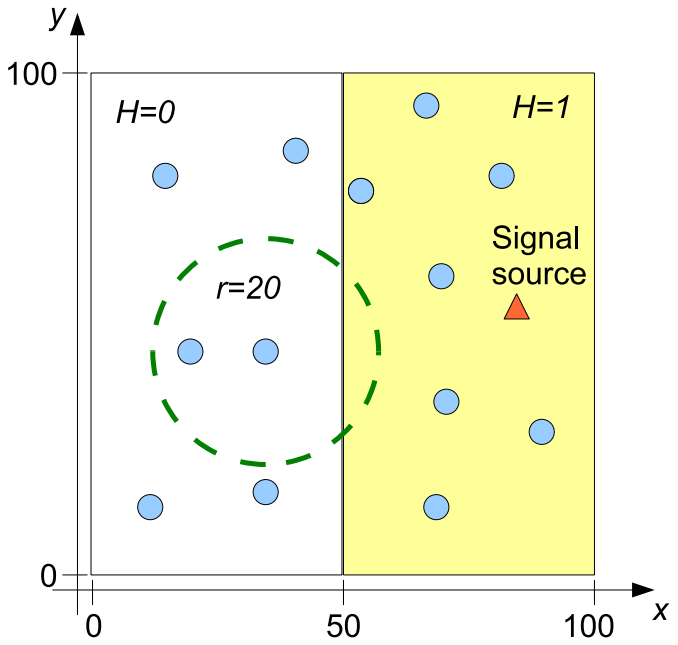


Figure 7.4: Large-scale network: $K = 50$ nodes randomly deployed in an area of 100×100 units; communication range $r = 20$ units. Nodes in the left sector ($x \leqslant 50$) are under $\mathcal{H} = 0$, nodes in the right sector ($x > 50$) are under $\mathcal{H} = 1$.

probability reaches exactly the performance of cooperative energy detection (CED), thus validating (7.35).

Note that in this example the threshold is like in the single-user case (7.6), therefore the achieved false-alarm probability is lower than the nominal rate α. In Fig. 7.6 the same setting is considered, but the threshold is now reset using (7.32), with $|\mathcal{C}| = 1 + |\mathcal{N}_2| = 3$. In this case, the system achieves the nominal false-alarm rate α asymptotically[2] in t, while the detection probability is significantly improved and still consistent with the CED bound (now computed using the cooperative threshold).

Fig. 7.7 shows the results obtained again in the small-scale network, but with different SNRs for the three nodes under $\mathcal{H} = 1$. In this case the performance of the NP-BP method converges to the CED curve (7.34) in the rightmost part of the plot, and it slightly outperforms CED for $\alpha \to 0$. The reason is that an additional advantage comes from the (negative) correlation with the nodes of the other cluster. Although this advantage is negligible in most cases, it becomes relevant when α is low and when the nodes in the cluster have different SNRs. To give insight into this fact, Fig. 7.8 compares the detection probability achieved (a) in the same scenario of the previous figure (two clusters, three nodes under $\mathcal{H} = 1$ and three under $\mathcal{H} = 0$), and (b) in the

2 In practice, for $t \approx 10$. Also note that expressions (7.6) and (7.32) are derived by applying the central limit theorem, and turn out to be slightly biased upwards for $\alpha \to 0$.

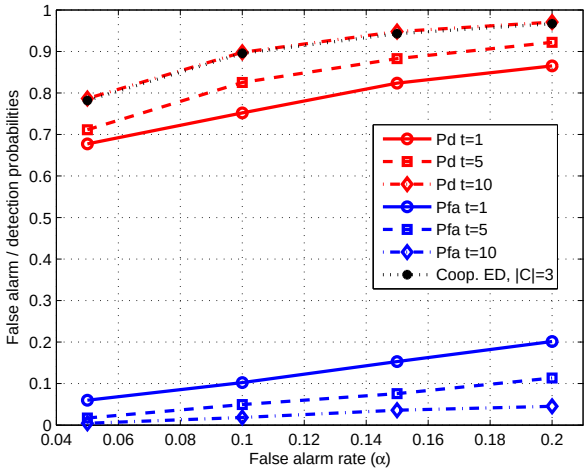


Figure 7.5: Simulation results for the small network of Fig. 7.3. SNR=$\{-5, -5, -5\}$ dB. 3 iterations. The reported P_d refers to node 2, P_{fa} to node 5.

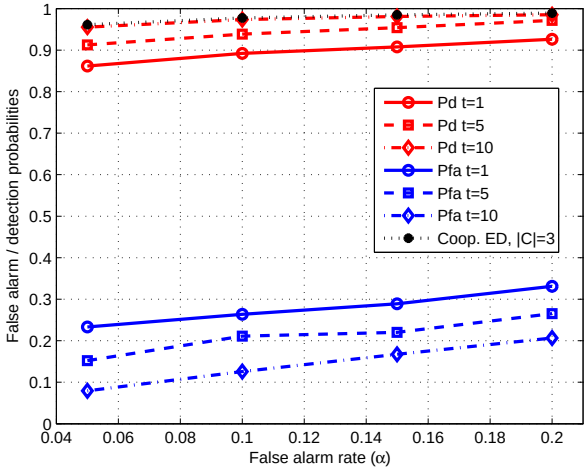


Figure 7.6: Simulation results for the small network of Fig. 7.3. SNR=$\{-5, -5, -5\}$ dB. 3 iterations. The reported P_d refers to node 2, P_{fa} to node 5. The threshold $\eta(\alpha)$ is set assuming $|\mathcal{C}| = 1 + |\mathcal{N}_2| = 3$ in (7.6).

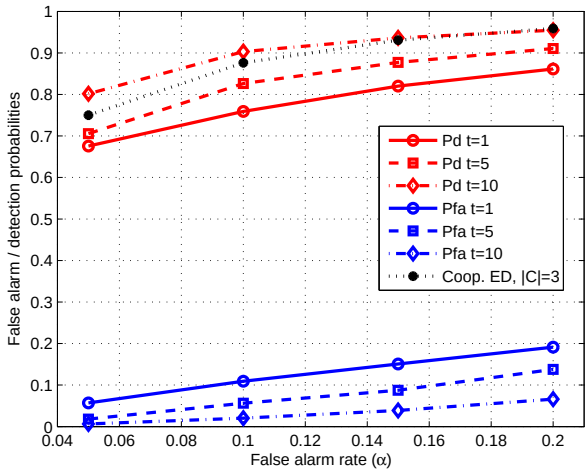


Figure 7.7: Simulation results for the small network of Fig. 7.3. SNR=$\{-4, -5, -7\}$ dB. 3 iterations.

first cluster alone. It can be observed that in the second case the CED curve is actually an upper bound, whereas in the presence of other nodes the bound can be attained and even passed. In all cases, however, the theoretical expression (7.34) provides useful information about the expected performance of the NP-BP method.

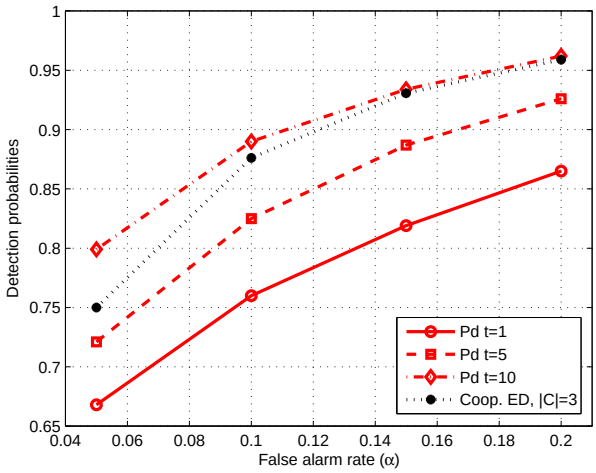
Next,the large-scale network of Fig. 7.4 is considered. For the nodes under $\mathcal{H} = 1$, the SNR is set randomly from a uniform distribution between -8 and -1 dB. The simulation results obtained in this scenario, shown in Fig. 7.9, confirm that NP-BP provides a substantial improvement over the performance of single-user detection. In this specific configuration, most of the improvement is obtained within the first 5 time slots. The threshold is set according to the single-user formula (7.6), hence the false-alarm probability is reduced as well.

Algorithm 7.1: Decentralized NP-BP algorithm

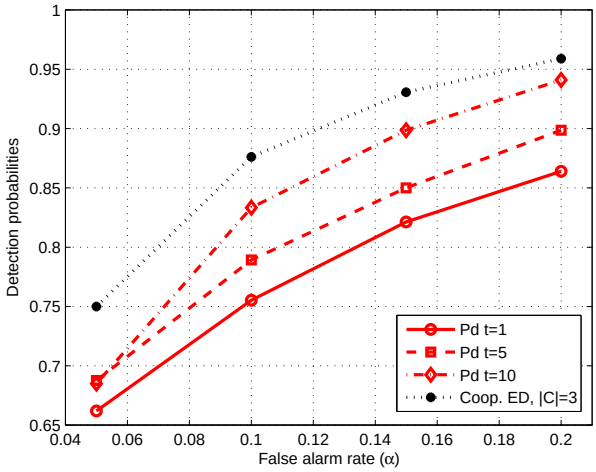
Input : Number of iterations L; false alarm rate α; max. number of time slots used for learning T_{max}; constant λ.

Output: Decisions $\{\hat{\mathcal{H}}_k^{(1)}, \ldots, \hat{\mathcal{H}}_k^{(S)}\}\ \forall k \in \{1, \ldots, K\}$.

for *time slot* $t = 1$ *to* S **do**
  Set $T = \min\{t, T_{max}\}$;
  for *nodes* $k \in \{1, \ldots, K\}$ **do**
    for *neighbors* $j \in \mathcal{N}_k$ **do**
      if $k < j$ **then**
        Update $\Delta_{kj}^{(t)}$ using (7.9);
      else
        $\Delta_{kj}^{(t)} = \Delta_{jk}^{(t)}$;
      end
    end
  end
  for *nodes* $k \in \{1, \ldots, K\}$ ***in parallel*** **do**
    Collect N received signal samples, $\boldsymbol{y}_k^{(t)}$;
    Compute $F_k^{(t)}$ from (7.7);
    Initialize and broadcast $M_{k\to j}^{(t,0)} = 0\ \forall j \in \mathcal{N}_k$;
    for *iteration* $l = 1$ *to* L **do**
      Receive incoming messages from neighbors: $M_{n\to k}^{(t,l-1)}$ $\forall n \in \mathcal{N}_k$;
      for *nodes* $j \in \mathcal{N}_k$ **do**
        Compute outgoing message $M_{k\to j}^{(t,l)}$ from (7.18) and send it to node j;
      end
    end
    Compute belief $\Lambda_k^{(t,L)}$ from (7.19);
    Make decision $\hat{\mathcal{H}}_k^{(t)}$ using (7.22);
  end
end



(a)



(b)

Figure 7.8: Detection performance comparison: single cluster vs. two clusters. (a) Configuration of Fig. 7.3, $K = 6$. (b) Only left cluster ($\mathcal{H} = 1$), $K = 3$. SNR=$\{-4, -5, -7\}$ dB.

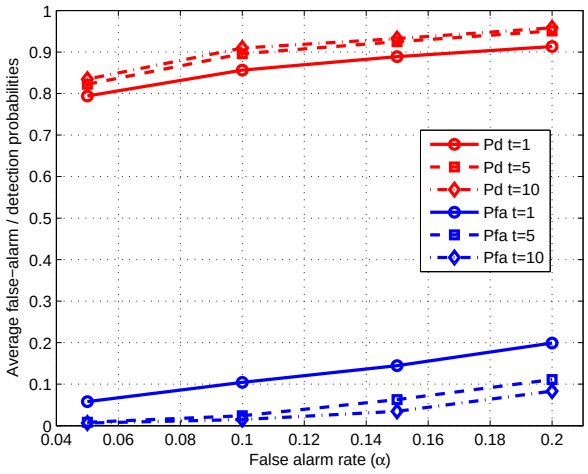


Figure 7.9: Simulation results for the large-scale network of Fig. 7.4. The SNRs of nodes under $\mathcal{H} = 1$ are chosen from a uniform distribution in $[-8, -1]$ dB. 3 iterations. The reported P_d and P_{fa} are averaged among all nodes under $\mathcal{H} = 1$ and $\mathcal{H} = 0$, respectively.

8

MULTI-SENSOR SIGNAL DETECTION WITH MALICIOUS SENSOR NODES

8.1 INTRODUCTION AND RELATED WORK

Cooperative spectrum sensing schemes have been extensively studied in the CR literature, e.g., [80, 168, 39]. Compared to single-user sensing, collaboration among different CR users provides robustness against fading environments and improved performance in terms of false alarm and missed detection. However, cooperative spectrum sensing is vulnerable to misbehaving nodes, which might report wrong sensing data, either because of malfunctioning radio, or intentionally as malicious users.

The problem of trustworthiness in cooperative spectrum sensing has been addressed in some recent works. For example, the method proposed in [165] assigns a "reputation value" to SUs. The reputation of each SU is updated at every time slot, based on the level of agreement of the considered SU and the majority of SUs. This method relies on the existence of a number of trusted SUs. If the reputation of a certain user is below a threshold, its sensing report is ignored. In [147] three types of attacks are considered: PHY layer attacks, MAC layer attacks, and combined PHY-MAC attacks. These attacks are addressed by an approach based on abnormality detection. In [172], another method based on no-regret learning is proposed. Game-theoretic strategies are investigated in [144]. In [146, 145, 148], a Bayesian inference method is proposed to detect malicious users, and an "onion-peeling" procedure is used to discard sensing reports from nodes considered as unreliable.

In this chapter a unified Bayesian framework is proposed to perform spectrum sensing and detection of misbehaving nodes in a combined fashion. A *statistical attack* model is adopted, whereby nodes behave maliciously with a certain probability. This type of attack is difficult to detect using traditional learning algorithms that assume a constant underlying pattern. Two possible attacks are considered: (a) *type-1 attackers* report that the channel is busy when it is free, thus preventing SU transmission and making the channel available for other, unauthorized, users, and (b) *type-0 attackers* report that the channel is free when it is busy, so as to increase the throughput of the secondary network by purposely violating the constraints of interference to the primary network. The proposed model has some similarities with the one developed in [146, 145, 148] ("CatchIt" algorithm), which defines attacks called "false-alarm" (similar to type-1) and "false-alarm & missed-detection" (a combination of type-1 and type-0). One difference is that in the CatchIt model attackers alter measured energy values (soft decision), whereas in the

present model they falsify reports (hard decision). A hard-decision model is motivated by two main reasons: (i) it can be applied to a broader class of detection techniques, not necessarily energy detection: for instance, eigenvalue-based detection [166], cyclostationary feature detection [30], etc.; (ii) it is suitable for distributed systems with limited bandwidth where the information is quantized to a single bit.

8.2 PROBLEM FORMULATION

The secondary network is composed of K users (SU) sensing the same channel at each time slot t. The presence of the PU is denoted by the binary variable $h^{(t)}$, where $h^{(t)} = 0$ means that PU is absent (free channel) at time t, and $h^{(t)} = 1$ means PU present (busy channel) at time t. Similar to [161, 163, 164], a hard decision reporting scheme is considered, where individual decisions from K independent nodes are gathered by a special node acting as the fusion center (FC). Let $u_k^{(t)}$ be a binary variable representing the true spectrum sensing decision of user k at time t, computed from observed statistics (e.g., the average energy of the received signal). The reported decision is then denoted by $y_k^{(t)}$. If the user is behaving maliciously, $y_k^{(t)}$ may differ from $u_k^{(t)}$.

In general, the probability $p(u_k^{(t)}|h^{(t)})$ can be expressed as

$$\begin{aligned} p(0|0) &= 1 - P_{fa}, \quad p(1|0) = P_{fa}, \\ p(0|1) &= 1 - P_d, \quad p(1|1) = P_d, \end{aligned} \tag{8.1}$$

where P_{fa} and P_d are, respectively, the probabilities of false alarm and detection, which depend on the specific detection technique adopted by SUs. For example, in the case of energy detection, P_{fa} and P_d are functions of the number of samples collected during time slot t, the noise power, the signal-to-noise ratio (SNR) of the primary signal, and the decision threshold.

Whether or not the reported decision $y_k^{(t)}$ differs from $u_k^{(t)}$ depends on user k's hidden variable r_k, that accounts for the *attack type/probability* of user k. In order to express the opposite behaviors of users of type-1 and type-0, variables r_k can be conveniently defined as follows[1]:

$$r_k = \begin{cases} = 0 & \text{user } k \text{ is honest,} \\ \in (0, 1] & \text{user } k \text{ is a type-1 attacker w.p. } r_k, \\ \in [-1, 0) & \text{user } k \text{ is a type-0 attacker w.p. } |r_k|. \end{cases} \tag{8.2}$$

1 Clearly, the mapping of user type/attack probability onto the variable r_k is purely conventional. Alternative representations are equivalent as long as they carry the same probabilistic meaning.

The reported decision can be expressed from $u_k^{(t)}$ and r_k as

$$y_k^{(t)} = \begin{cases} u_k^{(t)} \text{ w.p. } 1-r_k, 1 \text{ w.p. } r_k & \text{if } r_k > 0 \text{ and } u_k^{(t)} = 0, \\ u_k^{(t)} \text{ w.p. } 1-|r_k|, 0 \text{ w.p. } |r_k| & \text{if } r_k < 0 \text{ and } u_k^{(t)} = 1, \\ u_k^{(t)} & \text{otherwise.} \end{cases} \tag{8.3}$$

According to this set-up, type-1 attackers ($r_k > 0$) tend to report that the channel is occupied when it is not, thus preventing SUs from using the channel; conversely, type-0 attackers ($r_k < 0$) report that the channel is free when it is not, thus favoring the secondary network.

The following observations and remarks give more insight into the proposed attack model.

1. From the point of view of a malicious node k, r_k is a known probability value; from the point of view of the fusion center, r_k is an unknown random variable in the range $[-1, 1]$.

2. It is assumed that values r_k remain constant over T consecutive time slots, i.e., the attack strategy of malicious nodes does not change too quickly. Note that probabilistic attacks need a sufficient number of realizations to produce predictable effects on average, therefore very frequent changes of attack probabilities (e.g., every time slot) would not be meaningful even from the attackers' perspective.

3. The extreme cases $r_k = 1$ and $r_k = -1$ represent users being always malicious, i.e., sending constant reports ($y_k^{(t)} = 1 \ \forall t$ or $y_k^{(t)} = 0 \ \forall t$, respectively). Such "trivial" attacks, although effective in the very short term, can be easily identified after a few time slots. For this reason, the main focus is on statistical attacks with $0 < |r_k| < 1$.

4. This work does not investigate the problem from the point of view of malicious users, i.e., the strategy that may lead an attacker to choose a certain probability value $|r_k|$. It is only observed that, in order to balance the probability of being detected with the effectiveness of the attack, malicious nodes should intuitively choose attack probabilities in the range $0 < |r_k| < 1$. This is the rationale for the adopted model.

Sometimes, there may be a subset of $L < K$ *trusted* nodes ($r_k = 0$), known by the fusion center. The first L indices $(r_1, \ldots, r_L)$ are assigned to these nodes for notational convenience[2]. The vector notation $\boldsymbol{r} \triangleq [r_{L+1}, \ldots, r_K] \in [-1, 1]^{K-L}$ is introduced to denote the unknown attack probability values.

The goal of spectrum sensing is to infer the correct value of $h^{(t)}$ at every time slot t, in spite of possibly wrong reports sent by malicious nodes. Spectrum occupancy states of consecutive time slots, i.e., $h^{(t)}$ and $h^{(t+1)}$, are

2 Note that it is not assumed that the fusion center knows *all* the trusted nodes *a priori*: some nodes might be honest, but unknown at the fusion center. In a similar scenario, these nodes would be considered suspicious at the beginning (unknown r_k), but eventually the estimates of r_k would converge to zero.

assumed as independent[3], Considering T consecutive time slots, let $\boldsymbol{y}^{(t)} \triangleq [y_1^{(t)}, \ldots, y_K^{(t)}]^\dagger$ (where † means transpose) be the reports received in the time slot t from all K users, and by $\boldsymbol{Y} \triangleq [\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)}]$ the sequence of reports received over T consecutive slots. Similarly, define $\boldsymbol{U}$ as a matrix containing the $K \times T$ true sensing decisions, and by $\boldsymbol{h} = [h^{(1)}, \cdots, h^{(T)}]$ the vector of T unknown channel states to be estimated.

Given the above defined model, the following questions are of interest:

- What is the achievable spectrum sensing performance by a network with sensor nodes undergoing statistical attacks?
- Is there an efficient way to counteract such attacks, when the number of attackers, their identity, their attack type and probabilities, are unknown at the FC? In other words, can the FC "learn" the variables r_k while processing the spectrum sensing reports sent by SUs?
- Once nodes with intermediate values of r_k (i.e., $0 < |r_k| < 1$) are found, should they be completely excluded from the decision process or can their information be still exploited in some way?

The next sections will provide answers to these questions. First, the achievable performance under the proposed attack model is derived using different fusion rules. Then, it is shown that the values of r_k can actually be learned through an efficient message-passing algorithm, and, with a sufficient number of independent observations (time slots), the spectrum sensing performance gets very close to the performance achieved in the case of full knowledge of variables r_k. Finally, the answer to the third question is that all nodes with $0 \leqslant |r_k| \leqslant 1$ should be considered in the decision process in order to extract all the available information. In this sense, a joint estimation of variables r_k and $h^{(t)}$ turns out to be optimal.

8.3 IMPACT OF STATISTICAL ATTACKS ON SPECTRUM SENSING PERFORMANCE

This section analyzes how statistical attacks affect SU detection performance (i.e., the probability of error when estimating $h^{(t)}$) under the considered model. Results are first derived for the error probability of a single node, then cooperative detection is analyzed.

3 Other possible models for the evolution of $h^{(t)}$ are based on Markov chains, where $p(h^{(t+1)}|h^{(t)})$ takes into account the statistics of primary user traffic. In this work, a simpler uncorrelated model is adopted because (i) it is more general and (ii) the primary traffic statistics are usually unknown at the secondary users.

8.3.1 *Single Node Detection Performance*

Consider a generic node k and time slot t. To simplify notation, indices k and t can be dropped. The following quantities and functions are defined:

- $p_0 \triangleq p(h=0)$ and $p_1 \triangleq p(h=1)$ represent the prior distribution of h. Usually, $p_0 = p_1 = 1/2$ (no prior information available).
- $\widetilde{P_{fa}}(r)$ and $\widetilde{P_d}(r)$ are the effective false-alarm and detection probabilities of a malicious SU with attack probability r. Note that such probabilities are equal to P_{fa} and P_d if reports are unaltered.
- $\phi(u,h) \triangleq p(u|h)$ is the conditional probability of the true spectrum sensing outcome given the channel state. It is expressed directly from (8.1) and is reported in Table 8.1. Note that, for trusted nodes, $p(y|h) = \phi(y,h)$, where y is the report provided by the sensing node.

$\phi(u,h)$	$u=0$	$u=1$
$h=0$	$1-P_{fa}$	P_{fa}
$h=1$	$1-P_d$	P_d

Table 8.1: Values of $\phi(u,h)$.

- $\chi(u,r,y) \triangleq p(y|r,u)$ is the conditional probability of the report y given sensing outcome and attack probability. It can be derived from (8.3) and is charted in Table 8.2.

$\chi(u,r,y)$		$y=0$	$y=1$
$u=0$	$r>0$	$1-r$	r
	$r\leqslant 0$	1	0
$u=1$	$r\geqslant 0$	0	1
	$r<0$	$\lvert r\rvert$	$1-\lvert r\rvert$

Table 8.2: Values of $\chi(u,r,y)$.

- $\psi(h,r,y) \triangleq p(y|h,r) = \sum_{u=0}^{1} \chi(u,r,y)\phi(u,h)$ is the observation likelihood function and can be expressed from the quantities defined above.

Based on the above definitions, the false-alarm and detection probabilities $\widetilde{P_{fa}}(r)$ and $\widetilde{P_d}(r)$ for a single node can be readily computed as follows:

$$\widetilde{P_{fa}}(r) = \psi(0,r,1) \qquad \text{and} \qquad \widetilde{P_d}(r) = \psi(1,r,1). \tag{8.4}$$

The probabilities are summarized in Table 8.3. These values show that malicious users of type-1 increase the false-alarm rate, while improving the detection probability. On the contrary users of type-0 increase the probability

	$r > 0$	$r < 0$
$\widetilde{P_{fa}}(r)$	$P_{fa} + r(1 - P_{fa})$	$P_{fa}(1 - \lvert r\rvert)$
$\widetilde{P_d}(r)$	$P_d + r(1 - P_d)$	$P_d(1 - \lvert r\rvert)$

Table 8.3: False alarm and detection probabilities for nodes with $r \neq 0$.

of missed detection while reducing false alarms. However, the overall effect of malicious SUs is detrimental, as it is evident when analyzing the behavior of $\widetilde{P_{fa}}(r)$ and $\widetilde{P_d}(r)$ in the limit $P_d \to 1$, $P_{fa} \to 0$, i.e., in absence of sensing errors[4]. From Table 8.3 it follows that

$$r > 0: \quad \widetilde{P_{fa}}(r) \to r, \quad \widetilde{P_d}(r) \to 1 \tag{8.5}$$

$$r < 0: \quad \widetilde{P_{fa}}(r) \to 0, \quad \widetilde{P_d}(r) \to 1 - |r|. \tag{8.6}$$

Therefore, for malicious users, either the probability of false alarm or the probability of missed detection is always lower bounded by $|r| > 0$, even in absence of sensing errors.

Defining the probability of error $P_e \triangleq \Pr(y \neq h)$, it follows that (assuming uniform prior distribution of h)

$$P_e = p_0 \widetilde{P_{fa}}(r) + p_1(1 - \widetilde{P_d}(r)) = \frac{1 - (1 - |r|)(P_d - P_{fa})}{2}, \tag{8.7}$$

irrespective of the sign of r. Observe that the error probability, in the limit of no sensing errors ($P_{fa} \to 0$, $P_d \to 1$), is lower-bounded by $|r|/2$.

8.3.2 *Cooperative Detection Performance*

Now, the impact of statistical attacks on cooperative spectrum sensing (K collaborating users) is investigated. Here, cooperative sensing is formulated as a classic "simple hypothesis testing" problem, i.e., as a test of likelihood of the binary variable $h^{(t)}$, with $\boldsymbol{r}$ assumed as either perfectly known or unknown. In fact, there is no way of estimating $\boldsymbol{r}$ in a single time slot. Under this perspective, the reports received during a single time slot ($\boldsymbol{y}^{(t)}$) are sufficient statistics for spectrum sensing, since $h^{(t)}$ is assumed to be independent from previous and future time slots. Possible knowledge about PU traffic statistics can be incorporated into the prior distribution $p(h)$, with values of $p_0, p_1 \neq 1/2$.

Three possible scenarios can be distinguished depending on the amount of information available at the fusion center. For each scenario a maximum a posteriori (MAP)/ML statistical test is derived. The difference between MAP and ML is whether the prior distribution is assumed known for the variable to be estimated, in this case h. MAP/ML tests represent, in each scenario,

4 Note that such condition is ideal in the sense that no physical detector can achieve at the same time $P_d = 1$ and $P_{fa} = 0$. For this reason it is referred to as a "limit".

lower bounds for the achievable performance of any Bayesian/non-Bayesian detector.

1. *Ideal case.* If all systems parameters (P_{fa}, P_d, $\boldsymbol{r}$) are known at the FC, the MAP criterion is

$$\begin{aligned}\hat{h}_{MAP}(\boldsymbol{y}) &= \arg\max_{h\in\{0,1\}} \frac{p(\boldsymbol{y}|h,\boldsymbol{r})p(h|\boldsymbol{r})}{p(\boldsymbol{y}|\boldsymbol{r})} \\ &= \arg\max_{h\in\{0,1\}} p(h)\prod_{k=1}^{K}\frac{p(y_k|h,r_k)}{p(y_k|r_k)}. \end{aligned} \tag{8.8}$$

   Note that index t has been dropped since a single time slot is considered here. The denominator in (8.8) does not depend on h so it can be ignored. Therefore, a test can be defined in the form of a likelihood ratio:

$$\begin{aligned} S_{MAP-i}(\boldsymbol{y}) &\triangleq \frac{p_1}{p_0}\prod_{k=1}^{K}\frac{\psi(1,r_k,y_k)}{\psi(0,r_k,y_k)} \\ &= \frac{p_1}{p_0}\prod_{k:y_k=0}\frac{1-\widetilde{P_d}(r_k)}{1-\widetilde{P_{fa}}(r_k)}\prod_{k:y_k=1}\frac{\widetilde{P_d}(r_k)}{\widetilde{P_{fa}}(r_k)} \underset{\hat{h}=0}{\overset{\hat{h}=1}{\gtrless}} 1, \end{aligned} \tag{8.9}$$

   where the symbol ":" means "such that". The values of $\widetilde{P_d}(r_k)$ and $\widetilde{P_{fa}}(r_k)$ are different for each node k and, for trusted nodes, they are simply P_d and P_{fa}. The maximum-likelihood criterion $S_{ML,i}$ is equal to S_{MAP-i} without the prior term p_1/p_0.

2. *Unknown malicious nodes.* If the FC knows the values of P_{fa} and P_d of the nodes, but ignores the presence of malicious nodes, all reports are taken into account as if $r_k = 0\ \forall k$. The MAP criterion takes the following form:

$$S_{MAP-u}(\boldsymbol{y}) \triangleq \frac{p_1}{p_0}\prod_{k=1}^{K}\frac{\phi(y_k,1)}{\phi(y_k,0)} = \frac{p_1}{p_0}\left(\frac{1-P_d}{1-P_{fa}}\right)^{K_0}\left(\frac{P_d}{P_{fa}}\right)^{K_1} \underset{\hat{h}=0}{\overset{\hat{h}=1}{\gtrless}} 1, \tag{8.10}$$

   where K_0 and K_1 are, respectively, the number of zeros and ones in $\boldsymbol{y}$. Again, the corresponding ML criterion S_{ML-u} is obtained by replacing p_1/p_0 by 1 in the MAP expression.

   Alternatively, if the FC knows that there are malicious nodes with unknown attack probability, it can completely ignore the unreliable nodes and formulate the test using the trusted nodes only (if $L > 0$). In this case, the test (defined S_{MAP-r}) is similar to S_{MAP-u} except that instead of K, only the first L elements of $\boldsymbol{y}$ are considered.

3. *No available information.* If the FC has no knowledge of system parameters P_{fa} and P_d, the only possible test is a majority voting rule:

$$\hat{h}_{maj}(\boldsymbol{y}) \triangleq 1 \text{ if } K_1 > K_0, \quad 0 \text{ otherwise.} \tag{8.11}$$

The majority test can be interpreted as an *ML-u* test (8.10) assuming $P_{fa} = 1 - P_d < 0.5$. In other words, in absence of other information, the FC must weigh equally the possible types of errors (false alarm and missed detection); the assumption of error probabilities lower than 0.5 is reasonable for any properly designed detector. Also observe that if P_{fa} and P_d are unknown, the prior p_1/p_0 can not be taken into account in the decision; therefore, in this case MAP and ML are equivalent.

Then, let $P_{fa}(K) \triangleq \Pr(\hat{h} = 1|h = 0)$ and $P_d(K) \triangleq \Pr(\hat{h} = 1|h = 1)$ be the resulting probabilities of false alarm and of detection achieved by the above defined cooperative detection methods using K reports. These probabilities can be expressed, in general, as follows:

$$P_{fa}(K) = \sum_{\{\boldsymbol{y}:S(\boldsymbol{y})>1\}} \left[\prod_{k:y_k=0} (1 - \widetilde{P_{fa}}(r_k)) \prod_{k:y_k=1} \widetilde{P_{fa}}(r_k) \right] \tag{8.12}$$

$$P_d(K) = \sum_{\{\boldsymbol{y}:S(\boldsymbol{y})>1\}} \left[\prod_{k:y_k=0} (1 - \widetilde{P_d}(r_k)) \prod_{k:y_k=1} \widetilde{P_d}(r_k) \right]. \tag{8.13}$$

These formulas are difficult to compute in general because the condition $S(\boldsymbol{y}) > 1$ must be tested for each of the 2^K possible combinations of the vector $\boldsymbol{y}$. Nevertheless, tractable expressions can be derived in the case of K nodes with equal values of r_k.

Proposition 3. *Given K nodes performing cooperative spectrum sensing, with equal P_{fa}, P_d, and attack probability r, the probabilities of false alarm and detection are given by*

$$P_{fa}(K) = \sum_{l=\lceil K^* \rceil}^{K} \binom{K}{l} \widetilde{P_{fa}}(r)^l (1 - \widetilde{P_{fa}}(r))^{(K-l)} \tag{8.14}$$

$$P_d(K) = \sum_{l=\lceil K^* \rceil}^{K} \binom{K}{l} \widetilde{P_d}(r)^l (1 - \widetilde{P_d}(r))^{(K-l)}. \tag{8.15}$$

where the threshold K^* depends on the adopted fusion rule, namely:

$$K^*_{MAP-i} = \frac{\ln \frac{p_0}{p_1} + K \ln \frac{1-\widetilde{P_{fa}}(r)}{1-\widetilde{P_d}(r)}}{\ln \left(\frac{\widetilde{P_d}(r)}{\widetilde{P_{fa}}(r)} \cdot \frac{1-\widetilde{P_{fa}}(r)}{1-\widetilde{P_d}(r)} \right)} \tag{8.16}$$

$$K^*_{MAP-u} = \frac{\ln \frac{p_0}{p_1} + K \ln \frac{1-P_{fa}}{1-P_d}}{\ln \left(\frac{P_d}{P_{fa}} \cdot \frac{1-P_{fa}}{1-P_d} \right)} \tag{8.17}$$

$$K^*_{maj} = \frac{K}{2}. \tag{8.18}$$

$K^*_{ML,i}$ *and* $K^*_{ML,u}$ *are equal to* K^*_{MAP-i} *and* K^*_{MAP-u} *with* $\ln \frac{p_0}{p_1} = 0$.

Proof. From the expressions of $\widetilde{P_{fa}}$ and $\widetilde{P_d}$ in Table 8.3, it can be verified that for any $|r| \leqslant 1$, if $P_d > P_{fa}$, then $\widetilde{P_d}(r) > \widetilde{P_{fa}}(r)$. This property implies that in $S_{MAP-i}(\boldsymbol{y})$, as well as in $S_{MAP-u}(\boldsymbol{y})$, the only factors larger than one are those of the form $\widetilde{P_d}/\widetilde{P_{fa}}$, i.e., those corresponding to the K_1 "ones" of vector $\boldsymbol{y}$. Therefore, the event $S(\boldsymbol{y}) > 1$ is triggered when K_1 is larger than a certain value K^*. For S_{MAP-i}, assuming constant r for all nodes, the critical value of K_1 is the solution of the equation

$$\left(\frac{1-\widetilde{P_d}(r)}{1-\widetilde{P_{fa}}(r)}\right)^{K-K_1}\left(\frac{\widetilde{P_d}(r)}{\widetilde{P_{fa}}(r)}\right)^{K_1} > \frac{p_0}{p_1},$$

which after some algebraic manipulations yields $K_1 > K^*$ as in (8.16). For all $l > K^*$, there are $\binom{K}{l}$ configurations of vector $\boldsymbol{y}$ that satisfy the same condition. Since elements y_k are mutually independent conditioned on r and on $h = 0$, the probability of having l ones in $\boldsymbol{y}$ is $\widetilde{P_{fa}}(r)^l$, and the probability of the remaining $K-l$ zeros is $(1-\widetilde{P_{fa}}(r))^{(K-l)}$. This argument proves (8.14) for the case *MAP-i*.

For the other cases, the probability of every configuration of $\boldsymbol{y}$ remains the same (it depends on the attack model), but the threshold for $S(\boldsymbol{y}) > 1$ changes. Thus, (8.17) is solution of the equation $S_{MAP-u} > 1$, and (8.18) follows immediately from (8.11).

Similarly one can prove (8.15) where the only difference is that $\widetilde{P_{fa}}$ is replaced by $\widetilde{P_d}$. □

Remarks:

1. For $P_{fa} = 1 - P_d$ and $p_0 = p_1$, it can be verified that $K^*_{MAP-u} = K^*_{maj}$, which confirms the interpretation of the majority rule as a special case of the MAP/ML criterion with unknown prior distribution and sensing error probabilities.

2. Consider now the limiting case $P_d = 1$. In (8.10), it is easy to verify that $\frac{\phi(y_k,1)}{\phi(y_k,0)} = 0$ when $y_k = 0$. Hence, the *MAP-u* test becomes a "1-out-of-K" decision rule, i.e., $\hat{h}_{MAP-u} = 0$ if $\exists k : y_k = 0$, while $\hat{h}_{MAP-u} = 1$ only if all elements of vector $\boldsymbol{y}$ are equal to 1. Similarly, in (8.9) it can be observed that $\frac{\psi(1,r_k,y_k)}{\psi(0,r_k,y_k)} = 0$ for $y_k = 0$ and for any $r_k > 0$, therefore $\hat{h}_{MAP-i} = 0$ if $\exists k : y_k = 0, r_k > 0$.

   The conclusion is that, if $r_k > 0\ \forall k$, the *MAP-u* test converges to *MAP-i* as $P_d \to 1$: in other words, assuming all type-1 attackers and ideal detection probability, *the optimal test does not require knowledge of the attack probabilities*.

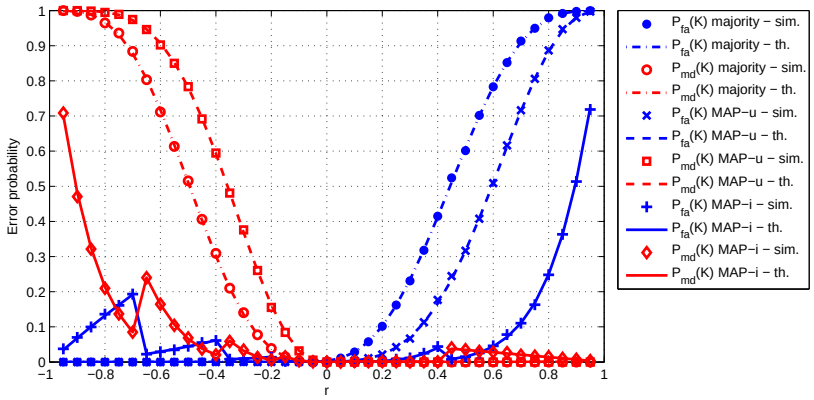


Figure 8.1: Global false alarm and missed detection probabilities ($P_{fa}(K)$ and $P_{md}(K)$) of collaborative spectrum sensing with $K = 7$ nodes with equal r, $P_{fa} = 0.1$, $P_d = 0.99$. Uniform prior distribution: $p_0 = p_1 = 0.5$. Theoretical (th.) vs. simulated (sim.) curves.

8.3.3 *Numerical Results*

The probabilities of false alarm, $P_{fa}(K)$, and missed detection, $P_{md}(K) \triangleq 1 - P_d(K)$, are plotted as a function of the attack probability r in Fig. 8.1. All nodes ($K = 7$) are assumed to have the same attack probability r and the same parameters $P_{fa} = 0.1$ and $P_d = 0.99$. The figure includes the three considered fusion rules: known parameters (*MAP-i*), unknown r (*MAP-u*), and majority voting. For each case the theoretical expressions of Proposition 1 are compared to simulation results.

As a general behavior, it can be noted that type-1 attacks tend to increase the global false-alarm probability, while type-0 attacks increase the missed detection probability. For all fusion rules, the worst performance is under attacks with large r ($r \to 1$ or $r \to -1$). However, when performing estimation of r over multiple time slots (see next section), such attacks can be easily identified, because a malicious node with $|r| = 1$ would report 1 or 0 all the time.

Comparing *MAP-u* and majority rule, it turns out that knowledge of P_{fa} and P_d does not necessarily help. If $P_{fa} > 1 - P_d$ (like in this case), the majority rule outperforms *MAP-u* in terms of $P_{md}(K)$, while it is inferior in terms of $P_{fa}(K)$. On the contrary, if $P_{fa} < 1 - P_d$, *MAP-u* provides better $P_{fa}(K)$ but higher $P_{md}(K)$. As previously noted, the two tests are equivalent for $P_{fa} = 1 - P_d$. In fact, for the values of P_{fa}, P_d chosen in this example, the statistical power of *MAP-u* and majority rule is equivalent; *MAP-u* is simply a majority rule with a modified threshold. These observations confirm that, in case of malicious nodes, estimating the attack probabilities r is essential for an accurate spectrum sensing.

The curve of *MAP-i* shows the achievable detection performance as a function of r, assuming perfect knowledge of r. It can be observed that the effect of attack probabilities is not monotonic, and exhibits "jumps". Interestingly, while the formulas of missed-detection/false-alarm probabilities of all considered detection strategies share the same structure (Eqs. 8.14, 8.15), only the *MAP-i* curves exhibit a non-monotonic behavior. The reason lies in the expression of K^*, which only for *MAP-i* varies with r (8.16). Then, discontinuities in the *MAP-i* performance curves correspond to values of r at which $\lceil K^*_{\text{MAP}-i}(r) \rceil$ changes. For example, referring to Fig. 8.1, a change of r from -0.65 to -0.75 triggers a new threshold $\lceil K^*_{\text{MAP}-i}(r) \rceil$, resulting in lower missed-detection probability and higher false-alarm probability.

The following sections introduce a new method to jointly perform spectrum sensing and estimating the attack probabilities, so as to achieve nearly ideal detection performance.

8.4 Proposed Algorithm

8.4.1 *Joint Spectrum Sensing and Estimation of Attack Probabilities*

It is now investigated how to improve the performance of spectrum sensing by exploiting a window of T consecutive time slots. If attack types/probabilities r_k remain constant over a certain number of slots, which is a reasonable assumption as discussed in Sec. 8.3, the values of $\boldsymbol{r} = [r_{L+1}, \ldots, r_K]$ can be estimated from the observed data $\boldsymbol{Y} = [\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)}]$ jointly with the state variables $\boldsymbol{h} = [h^{(1)}, \cdots, h^{(T)}]$. Adopting a Bayesian estimation approach, MAP or MMSE estimates of all variables of interest can be obtained from the marginal *a posteriori* probabilities, defined as

$$p(h^{(t)}|\boldsymbol{Y}) = \int d\{\boldsymbol{r}\} \sum_{\{\boldsymbol{h} \setminus h^{(t)}\}} \sum_{\{\boldsymbol{U}\}} p(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{U}|\boldsymbol{Y}), \tag{8.19}$$

$$p(r_k|\boldsymbol{Y}) = \int d\{\boldsymbol{r} \setminus r_k\} \sum_{\{\boldsymbol{h}\}} \sum_{\{\boldsymbol{U}\}} p(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{U}|\boldsymbol{Y}), \tag{8.20}$$

where

$$p(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{U}|\boldsymbol{Y}) \propto p(\boldsymbol{h}) p(\boldsymbol{r}) \prod_{t=1}^{T} \left[\prod_{k=1}^{L} \phi(y_k^{(t)}, h^{(t)}) \prod_{k=L+1}^{K} \chi(u_k^{(t)}, r_k, y_k^{(t)}) \phi(u_k^{(t)}, h^{(t)}) \right] \tag{8.21}$$

is the joint *a posteriori* probability given the observation matrix $\boldsymbol{Y}$, and $p(\boldsymbol{h})$ and $p(\boldsymbol{r})$ are the prior distributions of $\boldsymbol{h}$ and $\boldsymbol{r}$, which are mutually independent. The following notation has been introduced here:

- Probability functions with vector arguments represent joint probability of all elements in the said vector: e.g., $p(\boldsymbol{h}) = p(h^{(1)}, \ldots, h^{(T)})$.

- Sum/integration over a vector/matrix (in brackets) denotes sum/integration with respect to all elements of such vector/matrix over their respective supports: e.g., $\int \mathrm{d}\{\boldsymbol{r}\} = \int_{-1}^{1} \mathrm{d}r_{L+1} \cdots \int_{-1}^{1} \mathrm{d}r_K$; $\sum_{\{\boldsymbol{U}\}} = \sum_{u_{L+1}^{(1)}=0,1} \cdots \sum_{u_K^{(T)}=0,1}$
- Symbol $\setminus$ denotes set difference (treating vectors as sets): e.g., $\boldsymbol{r} \setminus r_k =$ all elements of $\boldsymbol{r}$ except r_k.

As discussed in Sec. 8.3, it is assumed $p(\boldsymbol{h}) = \prod_{t=1}^{T} p(h^{(t)})$ and $p(h^{(t)} = 0) = p(h^{(t)} = 1) = 1/2$ $\forall t$. Similarly, it is assumed that the fusion center does not have any prior information about attack probabilities of unreliable users, hence variables r_k are modeled as mutually independent, $p(\boldsymbol{r}) = \prod_{k=L+1}^{K} p(r_k)$, with $p(r_k)$ uniformly distributed in $[-1, 1]$ $\forall k$.

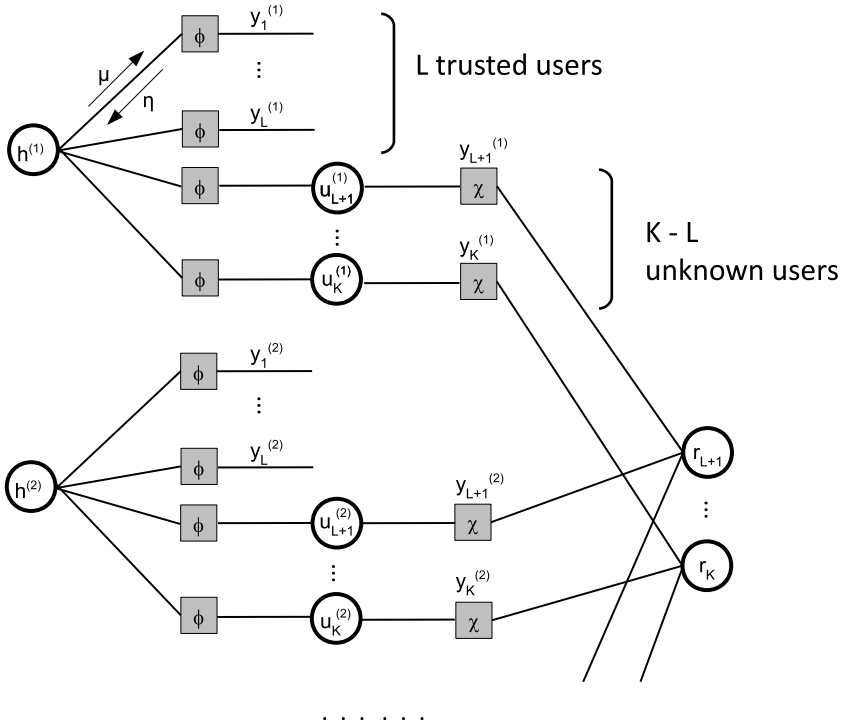


Figure 8.2: Factor graph of $p(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{U}|\boldsymbol{Y})$ truncated at $t = 2$. Each time slot t adds a branch to the graph, as indicated by the dots.

Direct computation of (8.21) may become extremely complex even for not-so-large numbers of users K and time slots T. In order to compute (or approximate) the marginal probabilities in a numerically efficient way, the BP algorithm (in the SPA version) [62] can be applied. A FG is constructed as in Fig. 8.2. At every iteration l, messages are updated according to the following rules.

- *Messages from variables to factors (μ)*

$$\mu^{(l)}_{h^{(t)}\to\phi_k^{(t)}}(h) \propto p(h^{(t)})\prod_{j\neq k}\eta^{(l-1)}_{\phi_j^{(t)}\to h^{(t)}}(h) \tag{8.22}$$

$$\mu^{(l)}_{u_k^{(t)}\to\chi_k^{(t)}}(u) = \eta^{(l)}_{\phi_k^{(t)}\to u_k^{(t)}}(u) \tag{8.23}$$

$$\mu^{(l)}_{u_k^{(t)}\to\phi_k^{(t)}}(u) = \eta^{(l)}_{\chi_k^{(t)}\to u_k^{(t)}}(u) \tag{8.24}$$

$$\mu^{(l)}_{r_k\to\chi_k^{(t)}}(r) \propto p(r_k)\prod_{s\neq t}\eta^{(l-1)}_{\chi_k^{(s)}\to r_k}(r). \tag{8.25}$$

- *Messages from factors to variables (η)*

$$\eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(h) \propto \sum_u \phi(u,h)\mu^{(l-1)}_{u_k^{(t)}\to\phi_k^{(t)}}(u) \tag{8.26}$$

$$\eta^{(l)}_{\phi_k^{(t)}\to u_k^{(t)}}(u) \propto \sum_h \phi(u,h)\mu^{(l-1)}_{h^{(t)}\to\phi_k^{(t)}}(h) \tag{8.27}$$

$$\eta^{(l)}_{\chi_k^{(t)}\to u_k^{(t)}}(u) \propto \int_{-1}^{1}\chi(u,r,y_k^{(t)})\mu^{(l-1)}_{r_k\to\chi_k^{(t)}}(r)\,dr \tag{8.28}$$

$$\eta^{(l)}_{\chi_k^{(t)}\to r_k}(r) \propto \sum_u \chi(u,r,y_k^{(t)})\mu^{(l-1)}_{u_k^{(t)}\to\chi_k^{(t)}}(u). \tag{8.29}$$

The marginal probabilities $p(h^{(t)}|\boldsymbol{Y})$ ($t = 1,\dots,T$) and $p(r_k|\boldsymbol{Y})$ ($k = L+1,\dots,K$) are thus approximated by *beliefs*. For every variable node, a belief is computed as the product of all messages entering the node, as follows:

$$b^{(l)}_{h^{(t)}}(h) \propto p(h^{(t)})\prod_{k=1}^{K}\eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(h) \tag{8.30}$$

$$b^{(l)}_{r_k}(r) \propto p(r_k)\prod_{t=1}^{T}\eta^{(l)}_{\chi_k^{(t)}\to r_k}(r). \tag{8.31}$$

All messages and beliefs are defined up to a normalization factor (hence the use of proportionality symbol, $\propto$), which has to be chosen such that each belief sums/integrates to 1. Note that normalization is not necessary at every step of the algorithm, but only when computing final beliefs. However, intermediate rescaling may be needed sometimes to avoid numerical problems (underflow).

As it is well-known, beliefs converge exactly to the true marginals probabilities only if the graph contains at most one cycle. However, BP is successfully used in many applications even in presence of cycles ("loopy BP"). In the model considered here (Fig. 8.2), cycles are generated by variable nodes r_k, which are connected to multiple (T) branches of the factor graph through factor nodes $\chi_k^{(t)}$. The reason is that variables r_k remain the same over T consecutive time slots, thus linking (in terms of statistical dependency) variables

at different time slots. Therefore, the number of loops in the graph increases with the number of time slots (T). On the other hand, however, increasing T is beneficial in that it adds more information to be used for inference of the hidden variables r_k.

Simulations (Sec. 8.4.3) and analytical results (Sec. 8.5) show that in the proposed application the performance of BP is not severely degraded by the presence of loops and, overall, the detection accuracy tends to improve as T grows.

8.4.2 *Spectrum Sensing with Sequential Estimation of Attack Probabilities*

Apart from the approximation introduced by BP, the proposed algorithm is theoretically optimal in the sense that, from a Bayesian point of view, the marginal a posteriori probabilities $p(h^{(t)}|\boldsymbol{Y})$ and $p(r_k|\boldsymbol{Y})$ contain the maximum amount of information that can possibly be inferred from the data $\boldsymbol{Y}$. However, joint estimation of $\boldsymbol{h}$ and $\boldsymbol{r}$ from the same data set (consisting of T time slots) involves a delay in the spectrum sensing decision, because the decision on $\boldsymbol{h}$, including $h^{(1)}$, is not available until all T report vectors $(\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)})$ are received.

To avoid this inconvenience, the proposed algorithm can be implemented in a slightly modified way. This alternative version works as follows: BP is run on a sliding window that consists of the current time slot plus $T-1$ previous slots, thus continuously updating estimates of r_k. These estimates are indicated as $\hat{r}_k$ (specifically, $\hat{r}_k^{\text{MMSE}}$ or $\hat{r}_k^{\text{MAP}}$, being, respectively, mean or maximum of $b_{r_k}(r)$). Computation of beliefs for $[h^{(T-1)}, \ldots, h^{(t-1)}]$ is now unnecessary; the decision on $h^{(t)}$ can then be simply made from the following MAP test

$$S_{\text{MAP-s}}(\boldsymbol{y}^{(t)}) \triangleq \frac{p_1}{p_0} \prod_{k=1}^{K} \frac{\psi(1, \hat{r}_k, y_k^{(t)})}{\psi(0, \hat{r}_k, y_k^{(t)})} = \frac{p_1}{p_0} \prod_{k:y_k=0} \frac{1-\widetilde{P_d}(\hat{r}_k)}{1-\widetilde{P_{fa}}(\hat{r}_k)} \prod_{k:y_k=1} \frac{\widetilde{P_d}(\hat{r}_k)}{\widetilde{P_{fa}}(\hat{r}_k)} \underset{\hat{h}^{(t)}=0}{\overset{\hat{h}^{(t)}=1}{\gtrless}} 1. \tag{8.32}$$

The above criterion can be interpreted as a *generalized likelihood ratio test* [154], where unknown parameters r_k are replaced by their estimates. This new algorithm is named *BP-sequential*, as opposed to the *BP-joint* presented in the previous section. Note that the test (8.32) produces the same result as direct computation of the mode of belief $b_{h^{(t)}}(h^{(t)})$ (8.30) because, in either way, $h^{(t)}$ is inferred using the values of $\boldsymbol{r}$ estimated by BP over the last T observations.

Joint and sequential BP are conceptually equivalent, but the latter is more efficient from a practical point of view because it does not involve any delay in the spectrum sensing process, while leveraging on the same amount of information (T time slots) to estimate $\boldsymbol{r}$. In addition, the sequential algorithm is

somewhat more flexible, because the estimation of $\boldsymbol{r}$ – based on BP, therefore numerically complex – is decoupled from that of $h^{(t)}$, which is now made through a simple test, (8.32). Therefore, in practical systems, estimation of $\boldsymbol{r}$ can be performed only from time to time (e.g., every 100 or 1000 time slots), to check for the presence of new malicious nodes or existing malicious nodes with changed attack probability, while using the last available values of $\hat{r}_k$ in the normal spectrum sensing routine. Observe that, for both methods, T must be chosen as a number of time slots over which $\boldsymbol{r}$ can be reasonably assumed as constant. As it will be shown in the next section, 15-20 time slots are enough to obtain a very accurate estimation of $\boldsymbol{r}$ (more precisely, accurate enough to nearly reach ML/MAP spectrum sensing performance).

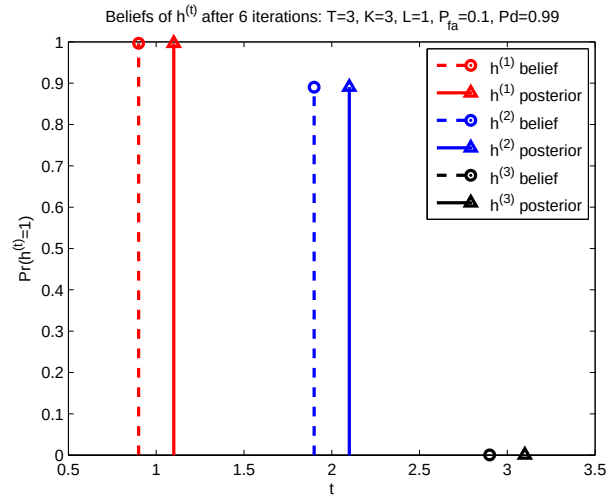
8.4.3 *Numerical Results*

Simulation results are now presented to investigate the performance of the proposed algorithm. In particular, the goal is to verify (i) whether the BP approximation is correct compared to the true marginal *a posteriori* probabilities, (ii) what is the performance gap between the proposed algorithm and ML/MAP detection with known attack probabilities, and (iii) how performance is affected by various system parameters, i.e., number of time slots, number of reliable users, and attack probability values. Then, in Sec. 8.5 a more formal analysis of the BP algorithm is presented, focusing on convergence and consistency issues.

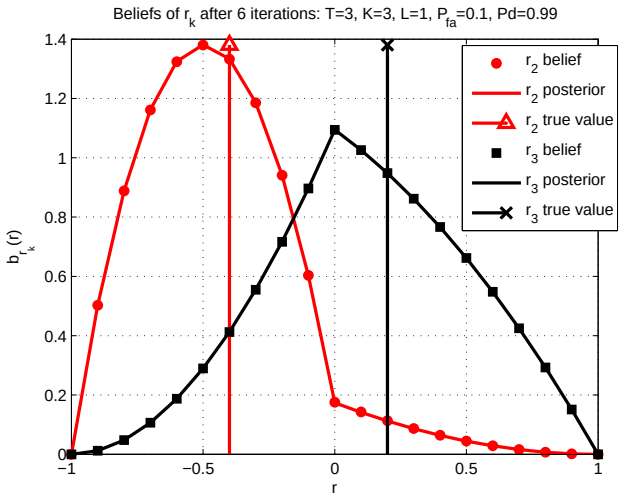
8.4.3.1 *Correctness of BP Solution*

The behavior of the proposed algorithm is first analyzed in a small graph ($T = 3$, $K = 3$, $L = 1$), where the exact marginal posterior distributions (8.19,8.20) can be computed analytically and compared to the beliefs obtained through BP. Results are shown in Fig. 8.3. It can be observed that beliefs of converged BP algorithms match perfectly with the posterior distributions, both for $h^{(t)}$ and r_k. Posterior distributions represent the optimal estimate in a Bayesian sense, so BP achieves its best possible performance. In this case, posteriors are also perfectly consistent with the true values of the unknown variables, resulting in a perfect estimation of $\boldsymbol{r}$ and $\boldsymbol{h}$, in spite of the small number of observations. Also note that the graph analyzed in this example contains three loops, therefore no theoretical results to the best of our knowledge guarantee exact convergence of beliefs (this aspect is discussed in more detail in Sec. 8.5). The empirical results observed here, confirmed by numerous other simulations, indicate that BP, in this specific application, is quite robust to the presence of loops.

Fig. 8.4 then shows an example of the evolution of beliefs as a function of iterations in a larger graph, with $T = 10$, $K = 4$, $L = 1$, which contains a large number of loops. In this case computation of the exact posteriors is prohibitive, therefore beliefs can be only compared against the true values

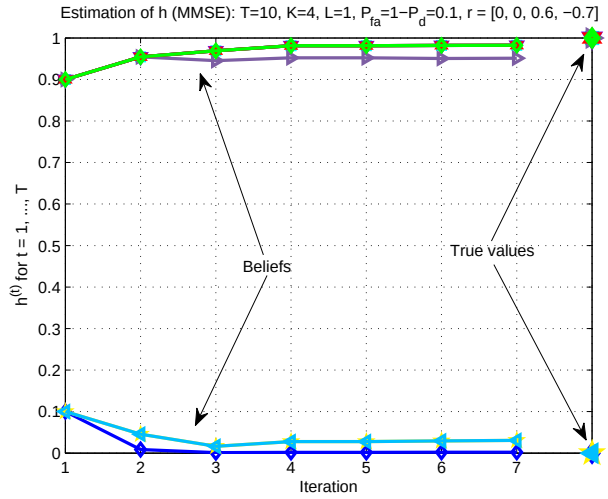


(a)

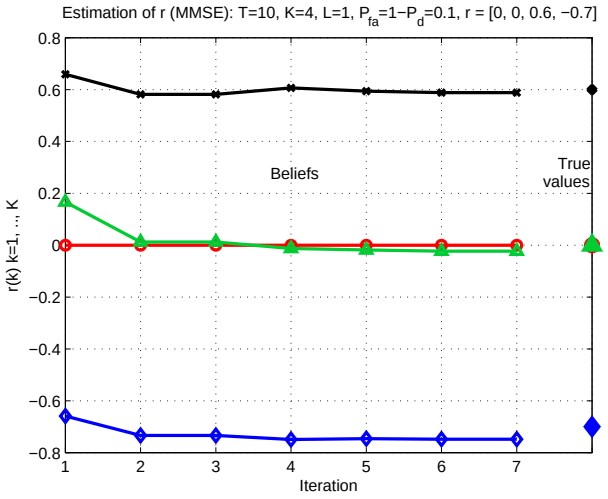


(b)

Figure 8.3: Beliefs at final iteration vs. marginal posterior probabilities of (a) $h^{(t)}$ and (b) r_k for a small graph ($T = 3$, $K = 3$, $L = 1$). True values: $h^{(1)} = 1$, $h^{(2)} = 1$, $h^{(3)} = 0$, $r_2 = -0.4$, $r_3 = 0.2$. BP results in exact convergence of beliefs in this case.



(a)



(b)

Figure 8.4: Evolution of MMSE estimates of beliefs (a) $b^{(l)}_{h^{(t)}}$ and (b) $b^{(l)}_{r_k}$ as a function of iterations (l). Each curve corresponds to a different user k (for variables r_k) or time index t (for variables $h^{(t)}$ – some curves are overlapping). True values are reported on the right side of each plot. Graph size: $T = 10$, $K = 4$, $L = 1$.

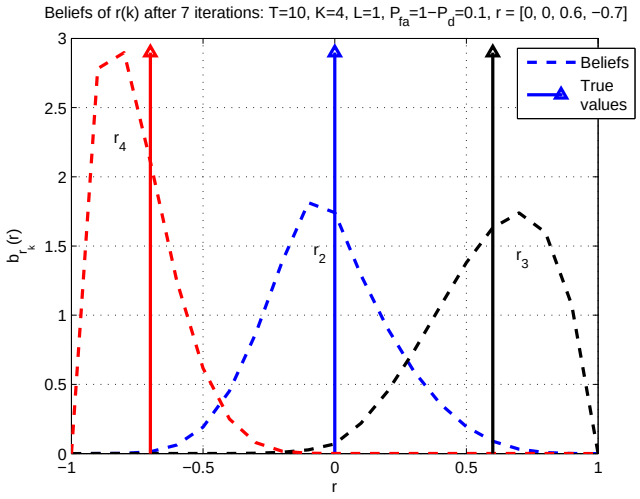


Figure 8.5: Beliefs $b_{r_k}(r)$ at final iteration vs. true value of r_k, in the same scenario as Fig. 8.4. Graph size: $T = 10$, $K = 4$, $L = 1$.

of r_k and $h^{(t)}$. It can be observed that (i) convergence is obtained after a few iterations (≈ 3), both for variables r and h, and (ii) the values of beliefs after convergence are very close to the true ones, despite the loops in the graph.

Fig. 8.5 shows in more details the distribution of the beliefs of variables r_k at convergence, referred to the same example as in the previous figure. All three beliefs turn out to be very close to their true values, both using MAP and MMSE estimation. Note that in simulations the support of variables r_k has been discretized with step of 0.1 and the trapezoidal integration method has been applied when computing (8.28).

8.4.3.2 *Spectrum Sensing Performance*

The performance of the proposed BP algorithm is now compared against the fusion rules presented in Sec. 8.3.2: the ideal MAP criterion *MAP-i* (8.9), assuming attack types/probabilities r_k as known exactly; the *MAP-u* criterion (8.10), with unknown r_k but known P_{fa}, P_d; the *MAP-r* criterion, defined like *MAP-u* but using the L reliable nodes only (if $L > 0$); and the majority voting rule (8.11). In all cases, the prior distribution of $h^{(t)}$ is assumed uniform $\forall t$, i.e., $p_0 = p_1 = 0.5$. The BP algorithm is implemented both in the "joint" version presented in Sec. 8.4.1 (*BP-joint*) and in the "sequential" version described in Sec. 8.4.2 (*BP-sequential*). In addition, a third possibility is considered, where nodes identified as malicious are completely discarded from the decision process. This method is named *BP-discarding*. In this case the test statistic is

$$S_{\text{MAP}-d}(\boldsymbol{y}) \triangleq \frac{p_1}{p_0} \prod_{k:r_k \leqslant \epsilon} \frac{\phi(y_k, 1)}{\phi(y_k, 0)}, \tag{8.33}$$

where ϵ is a sufficiently small value of attack probability, e.g., $\epsilon = 0.1$. This method is included in the comparison because almost all algorithms pro-

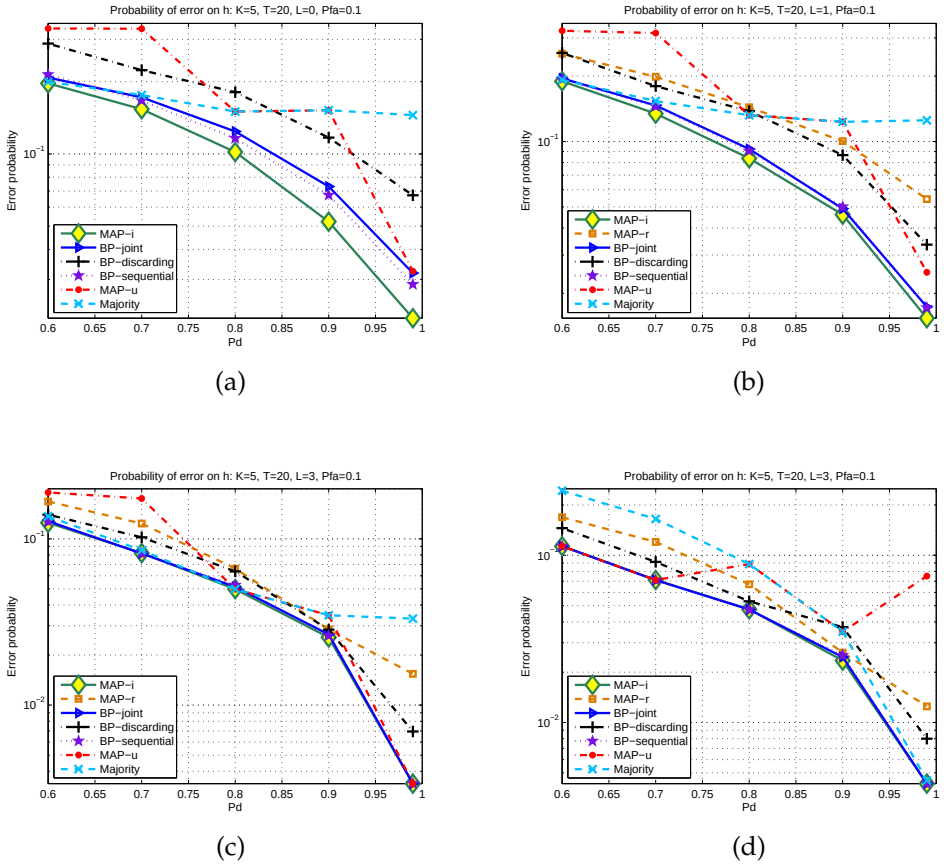


Figure 8.6: Probability of error on $h^{(t)}$ vs. P_d for different considered methods. $K = 5$, $T = 20$, $P_{fa} = 0.1$. (a) All users are malicious ($L = 0$); $\boldsymbol{r} = [0.1, 0.3, 0.4, 0.6, 0.2]$. (b) One user is trusted ($L = 1$); $\boldsymbol{r} = [0.3, 0.4, 0.6, 0.2]$. (c) Three users are trusted ($L = 3$); $\boldsymbol{r} = [0.6, 0.2]$. (d) Three users are trusted ($L = 3$); $\boldsymbol{r} = [-0.6, -0.2]$.

posed in the related literature (see, e.g., Sec. 8.1) are based on a detection-and-exclusion approach. For this reason, it is interesting to test whether the proposed algorithm performs better by using all sources of information, or only those that carry reliable information.

The results of this comparison are reported in Fig. 8.6, where the global error probability

$$P_e(K) \triangleq \Pr(\hat{h} \neq h) = p_0 P_{fa}(K) + p_1 P_{md}(K) \tag{8.34}$$

is plotted as a function of P_d, i.e., the detection capability of a single node, while keeping $P_{fa} = 0.1$ fixed. Since P_d generally increases with the SNR (e.g., using energy detection), the graph shown in Fig. 8.6 can be interpreted as a curve of detection performance vs. SNR, which is a typical performance figure in detection theory. The number of nodes is $K = 5$. Five values of

P_d are considered: [0.6, 0.7, 0.8, 0.9, 0.99] (the value $P_d = 1$ is excluded as not realistic in practice). Panel (a) refers to a scenario with all malicious users ($L = 0$), panel (b) to a scenario with one trusted user ($L = 1$), panels (c) and (d) to a case with $L = 3$ (the majority of nodes are trusted). In cases (a), (b), and (c), all attackers are of type-1 ($r_k > 0\ \forall k > L$), while in case (d) attackers are of type-0 ($r_k < 0\ \forall k > L$). Trusted nodes are sources of reliable information, therefore the performance of all considered methods tends to improve as L grows.

The curve of *MAP-i* is, as previously explained, the lower bound of achievable detection performance. Remarkably, the BP methods – both "joint" and "sequential" – attain practically the same performance as *MAP-i* when $L = 3$, and slightly lower performance when $L = 0$ or $L = 1$. Observe that the factor graph corresponding to this example contains a large number of loops, because $T = 20$. Therefore, loops do not seem to significantly affect the performance of BP in the proposed application. This intuition is supported by more rigorous arguments, presented in Sec. 8.5. It is worth noting, then, that *BP-joint* and *BP-sequential* are equivalent in terms of performance, as explained is Sec. 8.4.2.

The performance of the BP algorithms is primarily influenced by the number of time slots T, i.e., the amount of information available to estimate the values of variables r_k. As it is clear from these simulations, $T = 20$ is sufficient to achieve nearly-optimal performance (again, taking *MAP-i* as a benchmark). Given a certain value of T, the presence of reliable nodes can positively affect the detection performance: if there are one or more trusted nodes, the values of $h^{(t)}$ are estimated with increased accuracy, which in turn improves the estimation of r_k, and consequently the overall detection performance. It has been observed in other simulations that if T is not sufficiently large (e.g., $T < 10$), the performance of the BP algorithm tends to degrade especially in the region $P_d \approx 1$. This phenomenon can be explained by the fact that, as $P_d \to 1$, the data $\boldsymbol{Y}$ becomes less informative from the point of view of estimation of r_k.

Comparison between joint (or sequential) BP and BP-discarding reveals that exclusion of nodes identified as unreliable is not a good strategy. This is obviously true in the case of $L = 0$, but also for $L = 1$ and $L = 3$. This result shows that all available information must be taken into account, because even reports coming from nodes with high misbehavior probability contain some amount of useful information. The proposed BP algorithm, which jointly (or sequentially) estimates $h^{(t)}$ and r_k, makes it possible to extract such information by properly processing the data. As a consequence, the BP algorithm is robust even in the case of no honest users, whereas other solutions (e.g., [165]), presume the presence of a certain number of trusted users.

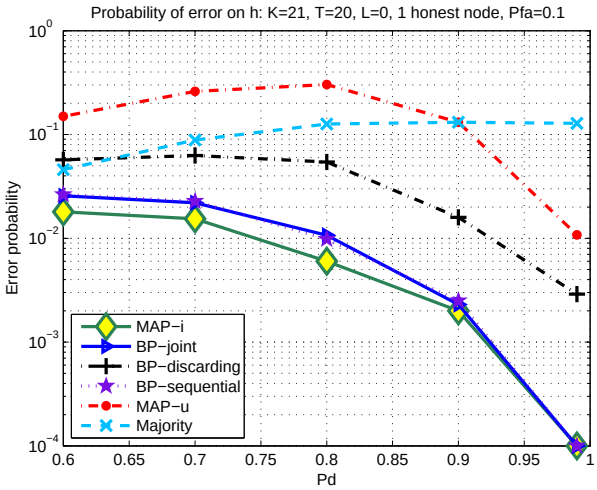
All the other fusion rules shown in Fig. 8.6 turn out to be suboptimal. In particular, the majority rule for type-1 attackers ($r_k > 0$: first three panels) provides poor performance in the high-P_d region, while for type-0 attackers

($r_k < 0$: last panel) it converges to *MAP-i* as $P_d \to 1$, but is suboptimal when P_d is low. The *MAP-u* curve exhibits a threshold behavior: in case of type-1 attackers, it converges to *MAP-i* when $P_d \to 1$ (as explained in Sec. 8.3.2, remark 2), but it results in very high error probability when P_d decreases; on the contrary, for type-0 attackers – panel (d) – it diverges from *MAP-i* in the high-P_d region and provides better performance when P_d is low. In all cases, *MAP-u* becomes equivalent to the majority rule when $P_d = 1 - P_{fa}$ ($= 0.9$ in this example), as observed in Sec. 8.3.2 (remark 1). To summarize, the majority rule and the *MAP-u* test are both suboptimal in some ranges of P_d (depending on malicious nodes' attack strategies), whereas the BP method proves to be robust over the entire P_d range.

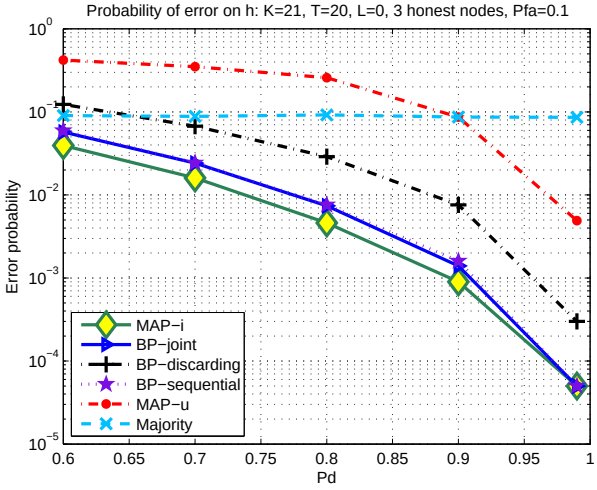
Finally, the curves of *MAP-r*, shown in panels (b), (c), and (d), represent the detection performance achieved by using the trusted node(s) only. *MAP-r* can be interpreted as an upper bound of the performance of the methods proposed in [165] ("reputation-based cooperative spectrum sensing", *RCSS*) and in [146] (*CatchIt*). Both algorithms, through different strategies, achieve identification of malicious nodes and exclusion of these nodes from the cooperative spectrum sensing process. Namely, *RCSS* compares the decision of each node against the majority of the other nodes' decisions, while *CatchIt* applies a more sophisticated Bayesian "onion-peeling" process. Two drawbacks are common to *RCSS* and *CatchIt*: (i) they do not work properly if a large number of nodes are malicious, and (ii) they involve decision thresholds (to decide whether to exclude nodes) which are set somewhat arbitrarily. Regardless of the above issues, the two algorithms would ideally retain the first L nodes (honest users) and discard all the other ones. As such, their best achievable performance is given by the *MAP-r* curve.

Thus, comparison between *MAP-r* and *BP-joint*/*BP-sequential* shows that the proposed BP methods significantly outperform *RCSS* and *CatchIt*. The gap is noticeable not only for $L = 1$, but also for $L = 3$, i.e., when trusted nodes are the majority. This result demonstrates that it is always beneficial to include all nodes in the decision process, instead of discarding nodes identified as malicious. In summary, the BP approach achieves high detection performance with any number of unreliable nodes, including the opposite cases $L = 0$ and $L > K/2$.

In Fig. 8.7, the same performance comparison of Fig. 8.6 is repeated for a larger-scale network, with unknown identification of reliable nodes. Specifically, the total number of nodes is $K = 21$. Two scenarios are considered: (a), where 1 node is reliable, and (b), where 3 nodes are reliable. In both cases, it is assumed that the fusion center does not know a priori the identity of honest users, i.e., $L = 0$. Results confirm the conclusions drawn for the previous example (Fig. 8.6), showing that: (i) the performance of the proposed sensing method remains close to the *MAP-i* bound when increasing the number of users, and (ii) prior knowledge of trusted users is not essential for the algorithm to work properly.



(a)



(b)

Figure 8.7: Probability of error on $h^{(t)}$ vs. P_d in a large-scale network ($K = 21$). Reliable users – 1 in panel (a), 3 in panel (b) – are unknown at the fusion center: $L = 0$. All malicious users are of type-1.

8.4.3.3 *Impact of System Parameters*

The impact of L and T on the the performance of the proposed algorithm is now investigated. The first panel of Fig. 8.8 shows the average error in estimation of $h^{(t)}$:

$$e_h \triangleq \frac{1}{T}\sum_{t=1}^{T} |\hat{h}^{(t)} - h^{(t)}|, \tag{8.35}$$

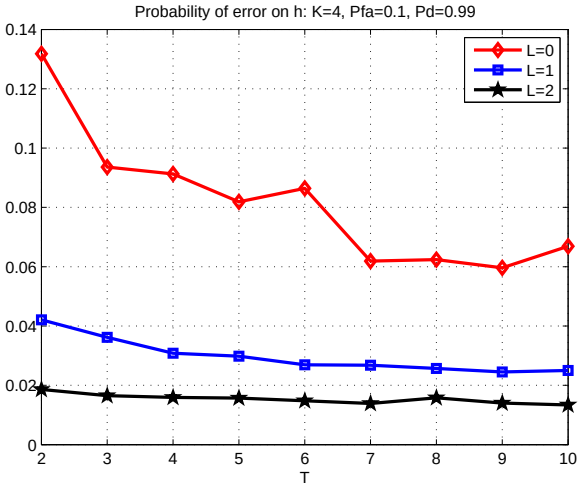
where $\hat{h} = \arg\max_{h=0,1} b_{h^{(t)}}(h)$. The second panel shows the root mean square error (RMSE) of the MMSE estimates of r_k:

$$e_r \triangleq \sqrt{\frac{1}{K-L}\sum_{k=L+1}^{K} |\hat{r}_k - r_k|^2}, \tag{8.36}$$

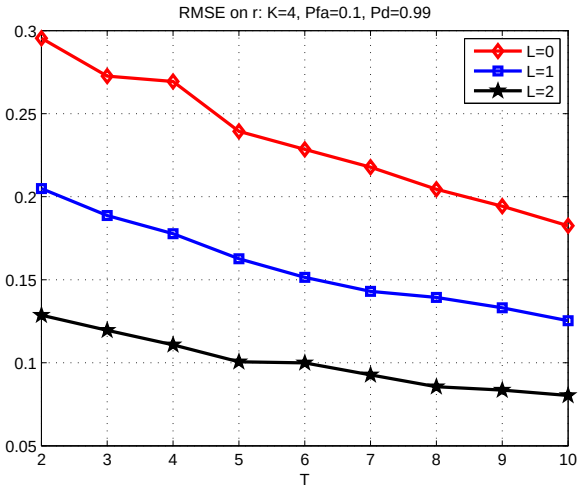
where $\hat{r}_k = \mathbb{E}_r b_{r_k}(r)$.

The first result is that L has a considerable impact on the performance of both r and h, because each trusted node introduces a factor that propagates an accurate belief of $h^{(t)}$ throughout the graph. Furthermore, the effect of any additional trusted node is multiplicative in P_{fa} and P_d, therefore the resulting false-alarm and detection probabilities scale exponentially with L. The other parameter, T, affects almost linearly the error on r_k. The impact of T on $h^{(t)}$ is evident especially for $L = 0$, where it is important to estimate the values r_k with high accuracy because the entire available information comes from malicious nodes and needs to be properly "filtered". On the contrary, for $L = 1$ the improvement as a function of T is more moderate, and for $L = 2$ the estimates of $h^{(t)}$ are already so accurate using the two trusted nodes, that the information coming from unreliable nodes has nearly no effect, hence additional slots are not needed. Note that the evolution of the estimation error as a function of T exhibits a generally decreasing trend, but is not necessarily monotonic, because loops in the graph (whose number increases with T) may render certain configurations slightly more favorable than others in spite of a lower amount of available information.

The following question is then addressed: what values of r_k are harder to estimate? Intuitively, it can be expected that low values of $|r_k|$ need more time slots to be estimated than values of $|r_k|$ close to 1. To this purpose, a scenario with $K = 3$ and $\boldsymbol{r} = [0.2\ 0.5\ 0.9]$ is considered, and the evolution of the RMSE vs. T is analyzed for each r_k, separately. Results are reported in Fig. 8.9. Apart from fluctuations due to the same reason explained before, a general trend can be observed: high attack probability values, e.g., $r_3 = 0.9$, can be estimated with higher accuracy (after a sufficient number of time slots) than low probability values, e.g., $r_1 = 0.2$. This result shows that attacks with high probability are effective in the short run, but can be effectively counteracted by applying the proposed BP algorithm, or possibly other learning/estimation methods. On the contrary, low-probability attacks



(a)



(b)

Figure 8.8: (a) Probability of error on $h^{(t)}$ and (b) RMSE of the MMSE estimate of r_k as a function of T and L. $K = 4$, $L = 1$, $\boldsymbol{r} = [0, 0.4, 0.6, 0.3]$.

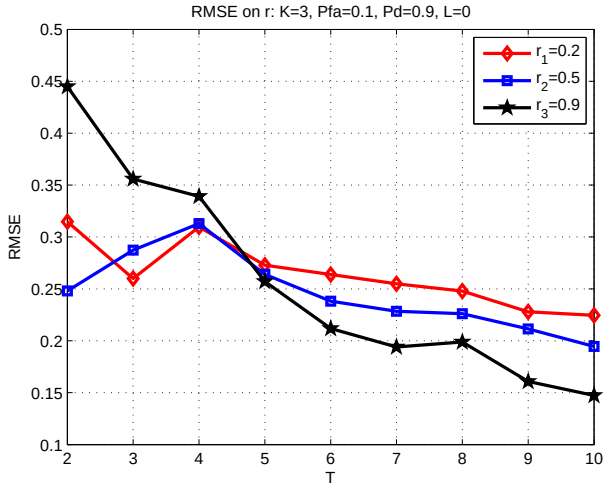


Figure 8.9: Probability of error on different r_k (RMSE of the MMSE estimator) as a function of T.

require more time to be detected with precision. Note that the sign of variables r_k, i.e., whether attackers are of type-1 or type-0, does not matter for estimation accuracy because the attack probability is given by the absolute value $|r_k|$. Simulations with negative values of r_k give identical results as in the positive case.

8.4.3.4 *Robustness to Time-Varying Attack Strategies*

Finally, in this section the performance of the proposed algorithm is investigated when an attacker's strategy is not constant but evolves over time. As an example, consider a scenario with $K = 2$ nodes, one of which being trusted ($L = 1$, i.e., $r_1^{(t)} = 0\ \forall t$) and the other one being malicious with an attack probability that gradually increases over time. Such behavior is modeled by a stepwise function for $r_2^{(t)}$, starting from an initial value $r_2^{(1)} = 0.1$ and augmenting by 0.2 every 5 time slots, up to $r_2^{(t)} = 0.9$ for $t \geqslant 21$. This way, the malicious user attempts to deceive the fusion center by inducing an underestimation of its attack probability. Observe that for an attack strategy to be meaningful in a probabilistic sense, the attack probability needs to remain constant at least for a certain number of time slots.

Then, the BP algorithm is applied under the above described conditions, with a graph extension of $T = 3$ time slots. As discussed in Sec. 8.3, T is a design parameter and, if there is a chance that attack probabilities vary frequently over time, it must not be too large, otherwise the detector would react slowly to changes in the values of r_k. The results of this experiment are illustrated in Fig. 8.10, where the true values of $r_2^{(t)}$ are compared against the estimates obtained at each time slot using BP beliefs. The BP algorithm proves to be robust to changes in the attack probability, and the estimates of

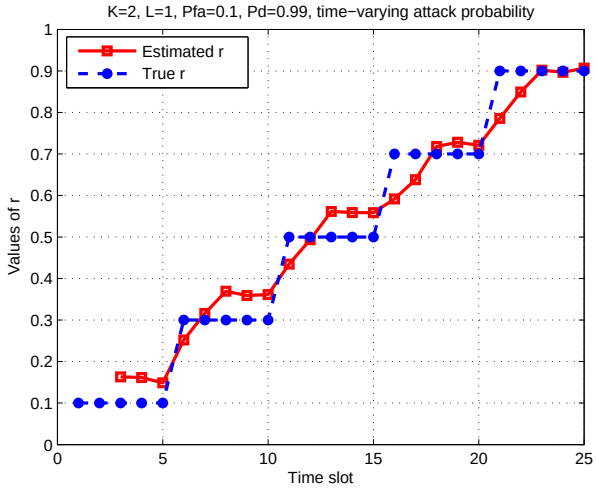


Figure 8.10: Estimation of a time-varying attack probability $r^{(t)}$ (stepwise function). A sliding window of length $T = 3$ time slots is used in the BP algorithm. Estimates of $r^{(t)}$ are averaged over 1000 Monte Carlo simulations.

r converge to the true values after a short transient period whose duration depends on T (the "memory" of the system). This example shows that the value $T = 3$ offers a balanced trade-off between reactivity and estimation accuracy.

8.5 THEORETICAL ANALYSIS OF THE BP ALGORITHM

This section aims at analyzing the dynamics of the proposed BP algorithm from a theoretical point of view. First, it is shown that, under certain limiting conditions, the BP algorithm is exact. Second, for general graphs, sufficient conditions are given for BP to converge to a unique value. Finally, it is shown that on average spectrum sensing through BP is consistent, with respect to the true value of $h^{(t)}$.

8.5.1 *Special Cases*

Based on well-known properties of BP, beliefs converge exactly to the true posteriors, i.e., $b^{(l)}_{h^{(t)}}|_{l\to\infty} = p(h^{(t)}|\boldsymbol{Y})$ and $b^{(l)}_{r_k}|_{l\to\infty} = p(r_k|\boldsymbol{Y})$ $\forall k, t$ in the following cases:

- *If* $T = 1$, *or if* $(T = 2) \wedge (K - L \leqslant 2)$. The reason is that BP is known to be exact in trees and in graphs with a unique loop [152]. It is immediate to verify that, when $T = 1$ (with any values of K, L) or when $T = 2, K - L = 1$, the factor graph is a tree, and when $T = 2, K - L = 2$ it has one loop only (see Fig. 8.11).

- *For any values of* K, T*, if there is at least one trusted user (*$L \geqslant 1$*) and in the limit of ideal sensing (*$P_{fa} \to 0, P_d \to 1$*).* In this case, $h^{(t)} \equiv y_1^{(t)} \equiv \ldots \equiv y_L^{(t)}\ \forall t$, hence variables $h^{(t)}$ are known exactly. At this point, the factor graph for variables r_k is an inference tree and therefore $b_{r_k}^{(l)}|_{l\to\infty} = p(r_k|\boldsymbol{Y}),\ \forall k$. Note that, in practice, the condition of no sensing errors is approached as L grows.

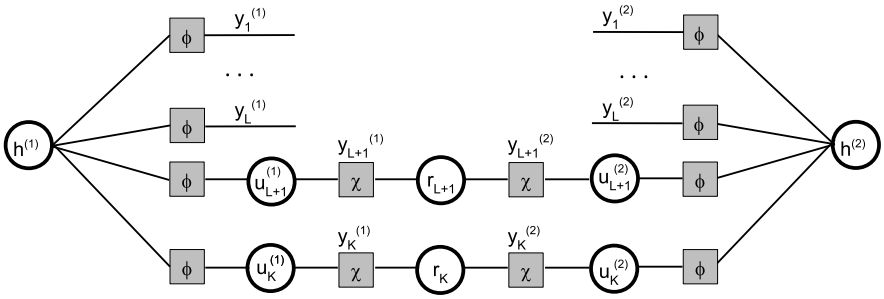


Figure 8.11: Factor graph for $T = 2, K - L = 2$.

8.5.2 *Convergence Analysis*

The asymptotic convergence ($l \to \infty$) of the BP algorithm in the proposed model is now analyzed for general graph.

Proposition 4. *BP is guaranteed to converge to a unique fixed point if the following conditions are simultaneously satisfied:*

$$\tanh\left[\frac{1}{4}\log\left(r^*\right)^2\right] < \frac{1}{T-1}, \tag{8.37}$$

and

$$\tanh\left(\frac{1}{4}\log p^*\right) < \frac{1}{K-L-1}, \tag{8.38}$$

where

$$r^* \triangleq \max\left\{\frac{r_{max}^+}{r_{min}^+}, \frac{|r_{max}^-|}{|r_{min}^-|}, \frac{1}{r_{min}^+|r_{min}^-|}\right\} \tag{8.39}$$

with $r_{max}^{\pm}, r_{min}^{\pm}$ *being respectively the largest and the smallest positive/negative elements of* $\boldsymbol{r}$*, and (assuming* $P_{fa} < 0.5$ *and* $P_d > 0.5$*)*

$$p^* \triangleq \frac{1-P_{fa}}{1-P_d}\cdot\frac{P_d}{P_{fa}}. \tag{8.40}$$

Proof. For a given factor graph, denote the factor nodes with capital letters (e.g., I, J) and the variable nodes with lower-case letters (e.g., i, j). Let ψ^I be

the probability function associated with factor I and x_i be the value taken by variable i. Denote the neighborhood set of a node x as $\mathcal{N}_x$. Now consider a subgraph defined by the path $J \to j \to I \to i$. Applying Theorem 3 in [82], it follows that BP converges if, for all such paths,

$$\max_{J \to j} \sum_{I \in \mathcal{N}_j \setminus J} \sum_{i \in \mathcal{N}_I \setminus j} N(\psi^I, i, j) < 1 \tag{8.41}$$

where $N(\psi^I, i, j)$ is defined as

$$N(\psi^I, i, j) = \sup_{\alpha \neq \alpha'} \sup_{\beta \neq \beta'} \sup_{\gamma, \gamma'} \tanh\left(\frac{1}{4} \log \frac{\psi^I_{\alpha\beta\gamma} \psi^I_{\alpha'\beta'\gamma'}}{\psi^I_{\alpha'\beta\gamma} \psi^I_{\alpha\beta'\gamma'}} \right) \tag{8.42}$$

with $\psi^I_{\alpha\beta\gamma} \triangleq \psi^I(x_i = \alpha, x_j = \beta, x_{I\setminus\{i,j\}} = \gamma)$. Here, $x_{I\setminus\{i,j\}} = \gamma$ is a shorthand notation for $x_k = \gamma \; \forall k \in \mathcal{N}_I \setminus \{i, j\}$.

Now, in the model of Fig. 8.2 the possible paths $J \to j$ are: $\chi_k^{(t)} \to u_k^{(t)}$, $\chi_k^{(t)} \to r_k$, $\phi_k^{(t)} \to h^{(t)}$, $\phi_k^{(t)} \to u_k^{(t)}$.

- For $\chi_k^{(t)} \to u_k^{(t)}$, neighbors are $\mathcal{N}_j \setminus J = \{\phi_k^{(t)}\}$ and $\mathcal{N}_I \setminus j = \{h^{(t)}\}$, and (8.41) reduces to

$$\max_{J \to j} N(\phi_k^{(t)}, h^{(t)}, u_k^{(t)}) < 1. \tag{8.43}$$

  Since $\tanh(x) < 1 \; \forall x$, this condition will always be satisfied. The same argument applies to $\phi_k^{(t)} \to u_k^{(t)}$.

- For $\chi_k^{(t)} \to r_k$, neighbors are $I \in \mathcal{N}_j \setminus J = \{\chi_k^{t'} : t' \neq t\}$ and $i \in \mathcal{N}_I \setminus j = \{u_k^{(t')}\}$. From (8.42), the following function needs to be evaluated:

$$\frac{\psi^I_{\alpha\beta\gamma} \psi^I_{\alpha'\beta'\gamma'}}{\psi^I_{\alpha'\beta\gamma} \psi^I_{\alpha\beta'\gamma'}} = \frac{\chi(u, r, y)\chi(u', r', y')}{\chi(u', r, y)\chi(u, r', y')} \triangleq f_\chi \tag{8.44}$$

  for $u \neq u'$, $r \neq r'$, and any $\{y, y'\}$. By analyzing all possible configurations (the ones that occur w.p. $\neq 0$), it can be verified that $\sup f_\chi = r^*$ as in (8.39). Since there are $T-1$ different factors I of the form $\{\chi_k^{t'} : t' \neq t\}$, and one variable i connected to each of them, the conclusion is (8.37).

- For $\phi_k^{(t)} \to h^{(t)}$, the neighbor set is given by $I \in \mathcal{N}_j \setminus J = \{\phi_l^{(t)} : l \neq k\}$, and $i \in \mathcal{N}_I \setminus j = \{u_l^{(t)}\}$. For factors with two variables, the function inside the logarithm in (8.42) reduces to (cf. [82], Eq. 48)

$$f_\phi \triangleq \frac{\phi(u, h)\phi(u', h')}{\phi(u', h)\phi(u, h')} \tag{8.45}$$

  for $u \neq u'$,$h \neq h'$. Note that, when $u = h$, $f_\phi = \frac{1-P_{fa}}{1-P_d} \cdot \frac{P_d}{P_{fa}}$, and when $u \neq h$, $f_\phi = \frac{1-P_d}{1-P_{fa}} \cdot \frac{P_{fa}}{P_d}$. Assuming $P_{fa} < 0.5$ and $P_d > 0.5$, then $\frac{1-P_{fa}}{1-P_d} \cdot \frac{P_d}{P_{fa}} > \frac{1-P_d}{1-P_{fa}} \cdot \frac{P_{fa}}{P_d}$ always, and therefore

$$\sup f_\phi = p^*. \tag{8.46}$$

  Since there are $K - L - 1$ such factors I, the result is (8.38).

□

An intuitive interpretation of conditions (8.37) and (8.38) is the following: convergence conditions become more and more restrictive as K and T increase, i.e., as the number of loops increases. Also, from (8.37) it turns out that convergence is guaranteed when all variables r_k have nearly the same value. For example, assume that all r_k are positive and that r_{max} is 2% larger than r_{min}: then, convergence is guaranteed for any $T \leqslant 101$. In contrast, if variables r_k are different in value and/or sign, e.g., $\boldsymbol{r} = [-0.1,\ 0.2]$, the convergence condition (8.37) becomes $T \leqslant 2$, i.e., it reduces to the trivial case discussed in Sec. 8.5.1. The second condition (8.38) is satisfied for larger values of K as P_d decreases while P_{fa} increases. In the extreme case of $P_d = P_{fa} = 0.5$, (8.38) is satisfied for any K. For low error probabilities, e.g., $P_{fa} = 0.1$ and $P_d = 0.99$, convergence is guaranteed only when $K \leqslant 2$, which again reduces to the special case considered in Sec. 8.5.1.

Remark: the results of Proposition 2 are only *sufficient* conditions for convergence. Simulations indicate that convergence actually occurs in a broader class of graphs. In fact, cases of non-convergence (oscillating beliefs) were never encountered in simulation. The problem of whether BP converges to the correct value is addressed next.

8.5.3 *Expectation of Beliefs*

In this section the evolution of the *expected values of beliefs* is analyzed, so as to give insight into the consistency of the BP algorithm. To this purpose, the "correct" element of the belief (e.g., $b(0)$ if $h^{(t)} = 0$) is compared against the "wrong" one (e.g., $b(1)$ if $h^{(t)} = 0$). More precisely, the following quantities are of interest:

$$\mathbb{E}_{y_k^{(t)}|h^{(t)}}[b^{(l)}_{h^{(t)}}(h^{(t)})] \text{ vs. } \mathbb{E}_{y_k^{(t)}|h^{(t)}}[b^{(l)}_{h^{(t)}}(\overline{h^{(t)}})], \tag{8.47}$$

where $\overline{h} \triangleq 1 - h$ denotes the complement of binary variable h. In order to achieve good performance, the first term should be consistently larger than the second one. To show that such property holds, an intermediate result needs to be proven first.

Lemma 2. If variables r_k are a priori independently and uniformly distributed, then the conditional probabilities of $y_k^{(t)}$ given $h^{(t)}$ are

$$\Pr(y_k = 1^{(t)}|h^{(t)} = 0) = \frac{1 + 2P_{fa}}{4} \tag{8.48}$$

$$\Pr(y_k = 1^{(t)}|h^{(t)} = 1) = \frac{1 + 2P_d}{4} \tag{8.49}$$

Proof. For $h^{(t)} = 0$,

$$\begin{aligned}
\Pr(y_k = 1^{(t)} | h^{(t)} = 0) &= \int_{-1}^{1} \left(p(r) \sum_{u=0,1} p(u|0) p(y|r,u) \right) dr \\
&= \int_{-1}^{1} \frac{1}{2} \left((1 - P_{fa}) \chi(0,r,1) + P_{fa} \chi(1,r,1) \right) dr \\
&= \frac{1}{2} \int_{0}^{1} \left((1 - P_{fa}) r + P_{fa} \right) dr + \frac{1}{2} \int_{-1}^{0} P_{fa} (1 - |r|) dr \\
&= \frac{1 + 2P_{fa}}{4},
\end{aligned}$$

which is (8.48). Similarly, (8.49) is found in the same way by letting $h^{(t)} = 1$. □

An important implication of Lemma 1 is that, given $P_{fa} < 0.5$ and $P_d > 0.5$, then $\Pr(y_k = 1^{(t)} | h^{(t)} = 0) < 0.5$ while $\Pr(y_k = 1^{(t)} | h^{(t)} = 1) > 0.5$. Furthermore, it is immediate to verify that $\mathbb{E}[y_k^{(t)} | h] = \Pr[y_k^{(t)} = 1 | h]$, therefore the probabilities (8.48) and (8.49) can be interpreted as expected values. Under this perspective, Lemma 1 means that, on average, reports are consistent although corrupted by malicious nodes.

Using this result, beliefs are shown to be consistent as well. Furthermore, beliefs provide an improvement in the confidence level of estimation of $h^{(t)}$ compared to reports $y_k^{(t)}$.

Proposition 5. Assume variables r_k, $h^{(t)}$ independent and with uniform prior distribution, and $P_d > 0.5, P_{fa} < 0.5$. Then for any user k, time slot t, and iteration number l, the expected values of beliefs of $h^{(t)}$ satisfy the relation

$$\mathbb{E}_{y_k^{(t)} | h^{(t)}} [b_{h^{(t)}}^{(l)} (h^{(t)})] > \mathbb{E}_{y_k^{(t)} | h^{(t)}} [b_{h^{(t)}}^{(l)} (\overline{h^{(t)}})]. \tag{8.50}$$

Proof. First recall that each belief (say $b_{h^{(t)}}$) is a scaled product of all incoming messages $\eta_{\phi_k^{(t)} \to h^{(t)}}$. Such incoming messages are conditionally independent given $h^{(t)}$. Therefore, to prove the proposition, it is sufficient to prove that, for all k,

$$\mathbb{E}_{y_k^{(t)} | h^{(t)}} \left[\eta_{\phi_k^{(t)} \to h^{(t)}}^{(l)} (h^{(t)}) \right] > \mathbb{E}_{y_k^{(t)} | h^{(t)}} \left[\eta_{\phi_k^{(t)} \to h^{(t)}}^{(l)} (\overline{h^{(t)}}) \right]. \tag{8.51}$$

Now, using (8.26), messages $\eta_{\phi_k^{(t)} \to h^{(t)}}^{(l)} (h^{(t)})$ can be written as

$$\eta_{\phi_k^{(t)} \to h^{(t)}}^{(l)} (h^{(t)}) \propto \phi(0, h^{(t)}) \mu_{u_k^{(t)} \to \phi_k^{(t)}}^{(l-1)} (0) + \phi(1, h^{(t)}) \mu_{u_k^{(t)} \to \phi_k^{(t)}}^{(l-1)} (1) \tag{8.52}$$

where, by applying (8.24) and (8.28),

$$\mu^{(l-1)}_{u_k^{(t)}\to\phi_k^{(t)}}(0) \propto \int_{-1}^{1}\chi(0,r,y_k^{(t)})\mu^{(l-2)}_{r_k\to\chi_k^{(t)}}(r)dr \tag{8.53}$$

$$\mu^{(l-1)}_{u_k^{(t)}\to\phi_k^{(t)}}(1) \propto \int_{-1}^{1}\chi(1,r,y_k^{(t)})\mu^{(l-2)}_{r_k\to\chi_k^{(t)}}(r)dr. \tag{8.54}$$

Note that the term $\mu^{(l-2)}_{r_k\to\chi_k^{(t)}}(r)$ is common to both (8.53) and (8.54), therefore after simple algebraic manipulations (8.52) can be rewritten as

$$\eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(h^{(t)}) \propto \int_{-1}^{1}\mu^{(l-2)}_{r_k\to\chi_k^{(t)}}(r)\left[\phi(0,h^{(t)})\chi(0,r,y_k^{(t)})+\phi(1,h^{(t)})\chi(1,r,y_k^{(t)})\right]dr. \tag{8.55}$$

For notational convenience, define $F_k^{(t,l)}(r) \triangleq \mu^{(l-2)}_{r_k\to\chi_k^{(t)}}(r)$. By applying recursively (8.25), (8.29), (8.23), (8.27) and (8.22), $F_k^{(t,l)}(r)$ can be expressed as

$$F_k^{(t,l)}(r) \propto \prod_{s\neq t}\left(\chi(0,r,y_k^{(s)})\left[(1-P_{fa})\prod_{j\neq k}\eta^{(l-4)}_{\phi_j^{(s)}\to h^{(s)}}(0)+(1-P_d)\prod_{j\neq k}\eta^{(l-4)}_{\phi_j^{(s)}\to h^{(s)}}(1)\right]\right.$$
$$\left.+\chi(1,r,y_k^{(s)})\left[P_{fa}\prod_{j\neq k}\eta^{(l-4)}_{\phi_j^{(s)}\to h^{(s)}}(0)+P_d\prod_{j\neq k}\eta^{(l-4)}_{\phi_j^{(s)}\to h^{(s)}}(1)\right]\right). \tag{8.56}$$

Note that is the above derivation constant terms are neglected in virtue of the assumption of uniform i.i.d. $p(r_k)$, $p(h^{(t)})$. Thus, from (8.55), the original messages $\eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(h^{(t)})$ can be written as

$$\eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(0) \propto \int_{-1}^{1}F_k^{(t,l)}(r)\left[(1-P_{fa})\chi(0,r,y_k^{(t)})+P_{fa}\chi(1,r,y_k^{(t)})\right]dr,$$

$$\eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(1) \propto \int_{-1}^{1}F_k^{(t,l)}(r)\left[(1-P_d)\chi(0,r,y_k^{(t)})+P_d\chi(1,r,y_k^{(t)})\right]dr,$$

i.e., as a function of messages of the same type, $\eta^{(l-4)}_{\phi_j^{(s)}\to h^{(s)}}$, sent four iterations earlier from nodes $\phi_j^{(s)}$ ($j\neq k, s\neq t$).

Then, let $\Delta \triangleq \eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(0)-\eta^{(l)}_{\phi_k^{(t)}\to h^{(t)}}(1)$. After some algebra Δ can be written as

$$\Delta \propto \int_{-1}^{1}F_k^{(t,l)}(r)(P_d-P_{fa})\left[\chi(0,r,y_k^{(t)})-\chi(1,r,y_k^{(t)})\right]dr. \tag{8.57}$$

In order to prove (8.51), the sign of the expected value of Δ must be studied for $h^{(t)}=0$ and for $h^{(t)}=1$. By linearity of expected value and integral,

$$\mathbb{E}_{y_k^{(t)}|h}[\Delta] \propto \int_{-1}^{1}F_k^{(t,l)}(r)(P_d-P_{fa})\mathbb{E}_{y_k^{(t)}|h}\left[\chi(0,r,y_k^{(t)})-\chi(1,r,y_k^{(t)})\right]dr.$$

(8.58)

Let $\chi_\Delta \triangleq \chi(0,r,y_k^{(t)}) - \chi(1,r,y_k^{(t)})$. The sign of $\mathbb{E}_{y_k^{(t)}|h}[\Delta]$ depends on the sign of $\mathbb{E}_{y_k^{(t)}|h}[\chi_\Delta]$, because $(P_d - P_{fa})$ is positive by assumption and $F_k^{(t,l)}(r) \geqslant 0$ $\forall r$ being a product of non-negative probability functions. Note that variables $y_k^{(s)}$ in $F_k^{(t,l)}(r)$ are independent from $y_k^{(t)}$. So, it follows that

$$\mathbb{E}_{y_k^{(t)}|h}[\chi(u,r,y_k^{(t)})] = \Pr(y_k^{(t)} = 0|h)\chi(u,r,0) + \Pr(y_k^{(t)} = 1|h)\chi(u,r,1), \tag{8.59}$$

hence

$$\begin{aligned}\mathbb{E}_{y_k^{(t)}|h}[\chi_\Delta] = &\Pr(y_k^{(t)} = 0|h)[\chi(0,r,0) - \chi(1,r,0)] \\ &+ \Pr(y_k^{(t)} = 1|h)[\chi(0,r,1) - \chi(1,r,1)].\end{aligned} \tag{8.60}$$

By definition of function χ Table 8.2 it readily found that, for both $r > 0$ and $r < 0$,

$$\chi(0,r,0) - \chi(1,r,0) = 1 - |r|, \tag{8.61}$$

$$\chi(0,r,1) - \chi(1,r,1) = |r| - 1, \tag{8.62}$$

therefore

$$\begin{aligned}\mathbb{E}_{y_k^{(t)}|h}[\chi_\Delta] &= (1-|r|)\left[\Pr(y_k^{(t)} = 0|h) - \Pr(y_k^{(t)} = 1|h)\right] \\ &= (1-|r|)\left[1 - 2\Pr(y_k^{(t)} = 1|h)\right]\end{aligned} \tag{8.63}$$

Finally, exploiting the expressions of $\Pr(y_k^{(t)} = 1|h)$ provided by Lemma 1, it follows that

$$\mathbb{E}_{y_k^{(t)}|h=0}[\chi_\Delta] = (1-|r|)\left(\frac{1}{2} - P_{fa}\right) > 0 \tag{8.64}$$

$$\mathbb{E}_{y_k^{(t)}|h=1}[\chi_\Delta] = (1-|r|)\left(\frac{1}{2} - P_d\right) < 0. \tag{8.65}$$

Thus, (8.51) is proved, which in turn implies (8.50). □

Remark. The message representation used in the above proof (Eqs. 8.57 and 8.57) allows for an intuitive interpretation of the BP algorithm. The term in squared brackets,
$\left[(1-P_{fa})\chi(0,r,y_k^{(t)}) + P_{fa}\chi(1,r,y_k^{(t)})\right]$ or $\left[(1-P_d)\chi(0,r,y_k^{(t)}) + P_d\chi(1,r,y_k^{(t)})\right]$,
representing the likelihood of the measurement $y_k^{(t)}$, is by itself sufficient to guarantee consistency. In fact, as a result of Lemma 1, reports $y_k^{(t)}$ are already consistent; Proposition 3 then proves that BP does not degrade such property irrespective of the number of loops. What is introduced by BP is the term $F_k^{(t,l)}(r)$, that can be interpreted as an estimate of the distribution of r obtained using to the information coming from other time slots $s \neq t$. Without BP, this distribution would be uniform. Thus, $F_k^{(t,l)}(r)$ can be interpreted as a "filter" of the observation likelihood function for different values of r.

9 REWEIGHTED BELIEF PROPAGATION

9.1 INTRODUCTION

In wireless communication, BP has found applications in many aspects of receiver design, including equalization, demapping, multi-user detection, multi-antenna detection, and decoding [156, 157, 28, 91]. Most notably in the latter application, the use of BP message passing over a suitable graphical model has led to practical decoding algorithms for powerful error-correcting codes, such as LDPC codes and turbo codes [153]. Recently, there has been an interest from the wireless communication community to extend BP to distributed problems, involving cooperation among multiple spatially separated wireless devices. BP inherently lends itself well to distributed implementation, and is thus a powerful, yet practical algorithm to perform cooperative estimation and detection. In fact, BP-based cooperative message passing algorithms have been applied to a wide variety of problems, including cooperative positioning [158, 49, 123], artificial intelligence [37], computer vision [127], cognitive radio [109, 65], link loss monitoring [74], network control [22], cooperative beamforming [90], and sensor networks [94]. Despite of its widespread use, BP is faced by a lack of convergence guarantees, as well as over-confidence of beliefs [152]. These problems have been partly addressed in [143], where a novel message passing algorithm was proposed, called tree-reweighted BP (TRW-BP). This algorithm has stronger convergence guarantees [118], and often gives better performance than BP. However, TRW-BP involves the selection of, and optimization over, so-called *edge appearance probabilities* (EAP). This makes TRW-BP difficult to implement in a network setting, as the choice of valid edge appearance probabilities involves a problem that is hard to solve in a distributed manner.

In this chapter, a novel algorithm is presented that combines the distributed nature of BP and the improved performance of TRW-BP in graphs with cycles. The new algorithm is called uniformly reweighted BP (URW-BP), as it collapses the edge appearance probabilities into a single scalar variable. As special cases, URW-BP reverts to BP under suitable choice of the variable, and corresponds to the optimal choice of edge appearance probabilities in TRW-BP for certain types of graphs. Although URW-BP is a straightforward modification of TRW-BP, it is a powerful inference algorithm in its own right, and deserves deeper investigation. The main contributions presented in the chapter are as follows:

- Introduction of URW-BP as a practical (and in some case equivalent) alternative to TRW-BP for distributed inference problems

- Interpretation of URW-BP as a multi-objective optimization problem, trading off single-variable entropy and correlation among variables, thus enabling URW-BP to mitigate overconfidence of beliefs in BP
- Extensive simulation results showing that URW-BP consistently outperforms BP, for centralized and distributed inference, for discrete and continuous variables, and for pairwise as well as higher-order interactions.

9.1.1 *Message Passing Algorithms*

Given a factor graph of $p(\mathbf{x}|\mathbf{y})$, a *message passing algorithm* can be run to determine approximations of the marginals $p(x_n|\mathbf{y})$, called beliefs, denoted by $b_n(x_n)$. Various message passing algorithms have been proposed, arguably the most popular of which is BP. Alternatives include naive mean field [142], TRW-BP [143], and generalized BP [162]. When the factor graph is cycle-free, BP and TRW-BP are equivalent and provide the exact marginals, i.e., $b_n(x_n) = p(x_n|\mathbf{y}), \forall n, x_n$. When the factor graph contains cycles, BP and TRW-BP are not guaranteed to yield the true marginal posteriors, or even to converge at all.

9.1.1.1 *Belief Propagation*

Assuming the variable X_n appears in the factor $\psi_{m,n}$, the BP message[1] from $\psi_{m,n}$ to X_n is given by

$$\mu_{\psi_{m,n}\to x_n}(x_n) = \sum_{x_m} \phi_m(x_m)\psi_{m,n}(x_m, x_n) \prod_{k\in\mathcal{N}_m\setminus\{n\}} \mu_{\psi_{m,k}\to x_m}(x_m), \tag{9.1}$$

where $k \in \mathcal{N}_m$ indicates that there exists a factor $\psi_{m,k}$. The beliefs are updated according to

$$b_n(x_n) \propto \phi_n(x_n) \prod_{m\in\mathcal{N}_n} \mu_{\psi_{m,n}\to x_n}(x_n). \tag{9.2}$$

Equations (9.1)–(9.2) are iterated until the beliefs converge.

9.1.1.2 *Tree-Reweighted Belief Propagation*

In TRW-BP, the message from $\psi_{m,n}$ to X_n is given by

$$\mu_{\psi_{m,n}\to x_n}(x_n) = \sum_{x_m} \phi_m(x_m)\psi_{m,n}^{1/\rho_{m,n}}(x_m, x_n)\mu_{\psi_{m,n}\to x_m}^{\rho_{m,n}-1}(x_m) \prod_{k\in\mathcal{N}_m\setminus\{n\}} \mu_{\psi_{m,k}\to x_m}^{\rho_{m,k}}(x_m), \tag{9.3}$$

1 For notational convenience, beliefs and messages will be expressed in terms of messages from factor vertices to variable vertices. Messages from variable vertices to factor vertices will not be considered, except in Section 9.2.3.

while the belief is given by

$$b_n(x_n) \propto \phi_n(x_n) \prod_{m \in \mathcal{N}_n} \mu_{\psi_{m,n} \to x_n}^{\rho_{m,n}}(x_n). \tag{9.4}$$

As in BP, (9.3)–(9.4) are iterated until the beliefs converge. TRW-BP, the values $\{\rho_{m,n}\}_{(m,n)\in\mathcal{E}}$ are the so-called edge appearance probabilities[2] (edge appearance probabilitys (EAPs)) of the factor vertices $\{\psi_{m,n}\}_{(m,n)\in\mathcal{E}}$.

The performance of TRW-BP depends on the choice of EAPs. Optimization over the EAPs (referred to here as *optimized TRW-BP*) is possible, but corresponds to an outer iterative loop, involving a high-dimensional optimization problem over $\{\rho_{m,n}\}_{(m,n)\in\mathcal{E}}$. Moreover, the set of valid EAPs is non-trivial: given a graph G and the set $\mathfrak{T}(G)$ of all possible spanning trees, a distribution over the spanning trees needs to be introduced: $0 \leqslant p(T) \leqslant 1$, for $T \in \mathfrak{T}(G)$. For a given distribution, the EAP of factor $\psi_{m,n}$ is then given by

$$\rho_{m,n} = \sum_{T \in \mathfrak{T}(G)} p(T) \times \mathbb{I}\{\text{vertex}\, \psi_{m,n} \in T\},$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. Note that when G is a tree, every factor vertex $\psi_{m,n}$ appears in every spanning tree, so that $\rho_{m,n} = 1$, $\forall (m,n) \in \mathcal{E}$. When the graph G contains cycles, there must be at least one $\rho_{m,n} < 1$.

BP can now be interpreted as a modification of TRW-BP, where $\rho_{m,n} = 1$, $\forall (m,n) \in \mathcal{E}$, irrespective of the structure of G. Note that this choice of EAPs is not valid for a graph with cycles.

9.2 UNIFORMLY REWEIGHTED BELIEF PROPAGATION

9.2.1 *Description*

In order to combine the simplicity and distributed nature of BP with the improved performance of TRW-BP in graphs with cycles, a novel approximate inference algorithm, called uniformly reweighted belief propagation (URW-BP), is proposed. In URW-BP, the EAPs in (9.3)–(9.4) are all equal (and denoted by ρ), so their optimization is a scalar optimization problem. The following sections will connect BP, TRW-BP, and URW-BP in a variational interpretation, comment on the conditions under which URW-BP and optimized TRW-BP coincide, and cast URW-BP as a solution to a multi-objection optimization problem.

2 The terminology of *edge* appearance comes from the Markov random field representation of $p(\mathbf{x}|\mathbf{y})$, where the factors $\psi_{m,n}$ are represented by edges.

9.2.2 *Variational Interpretation*

9.2.2.1 *General Formulation*

Given any distribution $b(\mathbf{x})$, the Kullback-Leibler divergence (KLD) [23] between $b(\mathbf{x})$ and $p(\mathbf{x}|\mathbf{y})$ is given by

$$\mathrm{KL}\left(b\|p\right) = \sum_{\mathbf{x}} b(\mathbf{x}) \log \frac{b(\mathbf{x})}{p(\mathbf{x}|\mathbf{y})} \geqslant 0.$$

By inserting (5.8), and perform some straightforward manipulations, this inequality can be expressed as

$$\log p(\mathbf{y}) \geqslant \mathcal{H}(b) + \underbrace{\sum_{n=1}^{N} \sum_{x_n} b_n(x_n) \log \phi_n(x_n) + \sum_{(m,n)\in\mathcal{E}} \sum_{x_m,x_n} b_{m,n}(x_m,x_n) \log \psi_{m,n}(x_m,x_n)}_{\triangleq \chi(b)}, \tag{9.5}$$

where $\mathcal{H}(b)$ denotes the entropy of the distribution $b(\mathbf{x})$, i.e.,

$$\mathcal{H}(b) = -\sum_{\mathbf{x}} b(\mathbf{x}) \log b(\mathbf{x}). \tag{9.6}$$

Note that (9.5) is valid with equality if and only if $b(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$, $\forall \mathbf{x}$. Hence,

$$\log p(\mathbf{y}) = \max_{b\in\mathbb{M}(G)} \left\{\mathcal{H}(b) + \chi(b)\right\}, \tag{9.7}$$

where $\mathbb{M}(G)$ is the so-called *marginal polytope* corresponding to the factor graph G, which is the set of marginal distributions $\{b_n(x_n)\}_{n=1}^N$ and $\{b_{m,n}(x_m,x_n)\}_{(m,n)\in\mathcal{E}}$ that can be related to a global distribution $b(\mathbf{x})$ that factorizes according to the same factor graph G. The formulation (9.7) implies that if one can solve the optimization problem, then the solution is $b(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$, with corresponding maximum equal to $\log p(\mathbf{y})$. The optimization problem turns out to be convex, but intractable: unless G is a tree, computing $\mathcal{H}(b)$ is NP-complete and the number of constraints to describe $\mathbb{M}(G)$ is exponential in the size of G. BP, TRW-BP, and URW-BP can be interpreted as approximate methods to solve (9.7).

9.2.2.2 *Belief Propagation*

In BP, the set $\mathbb{M}(G)$ is outer-bounded by the convex set $\mathbb{L}(G)$, which is the set of marginal beliefs $\{b_n(x_n)\}_{n=1}^N$ and $\{b_{m,n}(x_m,x_n)\}_{(m,n)\in\mathcal{E}}$, that are mutually consistent, but need not be consistent with any global belief $b(\mathbf{x})$. It can be shown that $\mathbb{L}(G)$ is a convex set, and that when G is a tree, $\mathbb{M}(G) = \mathbb{L}(G)$ [142]. Secondly, the entropy in (9.7) is replaced by the so-called Bethe entropy

$$\mathcal{H}_{\text{Bethe}}(b) = \sum_{n=1}^{N} \mathcal{H}(b_n) - \sum_{(m,n)\in\mathcal{E}} \mathcal{I}(b_{m,n}), \tag{9.8}$$

where $\mathcal{I}(b_{m,n})$ is the mutual information between x_m and x_n:

$$\mathcal{I}(b_{m,n}) = \sum_{x_m, x_n} b_{m,n}(x_m, x_n) \log \frac{b_{m,n}(x_m, x_n)}{b_m(x_m) b_n(x_n)}.$$

It is shown in [162] that BP (9.1)–(9.2) is an iterative method to find of a stationary point of the Lagrangian.

9.2.2.3 *Tree-Reweighted Belief Propagation*

In TRW-BP, the set $\mathbb{M}(G)$ is also outer-bounded by the convex set $\mathbb{L}(G)$. For a fixed set of valid EAPs, the entropy $\mathcal{H}(b)$ is replaced by a convex *upper* bound

$$\mathcal{H}(b|\boldsymbol{\rho}) = \sum_{n=1}^{N} \mathcal{H}(b_n) - \sum_{(m,n)\in\mathcal{E}} \rho_{m,n} \mathcal{I}(b_{m,n}),$$

where $\boldsymbol{\rho} \triangleq [\rho_{m,n}]_{(m,n)\in\mathcal{E}}$. Note again that BP corresponds to setting $\boldsymbol{\rho} = \mathbf{1}$ in TRW-BP. For a fixed $\boldsymbol{\rho}$, the stationary points of the Lagrangian lead to (9.3)–(9.4), following a similar line of reasoning as in [162].

9.2.2.4 *Uniformly Reweighted Belief Propagation*

In URW-BP, similar to BP and TRW-BP, the set $\mathbb{M}(G)$ is outer-bounded by the set $\mathbb{L}(G)$. For a fixed value of $\rho \in \mathbb{R}$, the entropy $\mathcal{H}(b)$ is approximated by

$$\mathcal{H}(b|\rho) = \sum_{n=1}^{N} \mathcal{H}(b_n) - \sum_{(m,n)\in\mathcal{E}} \rho \mathcal{I}(b_{m,n}). \tag{9.9}$$

Again, BP is found by setting $\rho = 1$. Stationary points of the Lagrangian lead to the TRW-BP equations given in (9.3)–(9.4), with $\rho_{m,n} = \rho$, $\forall (m,n) \in \mathcal{E}$.

9.2.3 *Extension of TRW-BP and URW-BP Beyond Pairwise Interactions*

In the original formulation of TRW-BP, only pairwise interactions were considered, with a corresponding Markov random field where variables are vertices and edges are pairwise potentials. It was suggested in [143] that the extension to higher-order interactions requires a graphical model that is a hypergraph. Based on factor graphs, TRW-BP equations have been derived for arbitrary interactions [160]. In particular, given a joint a posteriori distribution of the form

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^{N} \phi_n(x_n) \prod_{l=1}^{L} \psi_l(x_{C_l}), \tag{9.10}$$

where $x_{C_l} \subseteq \mathbf{x}$, the message from factor vertex ψ_l to variable vertex X_n is given by

$$\mu_{\psi_l \to x_n}(x_n) = \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \times \prod_{m \in C_l \setminus \{n\}} \phi_m(x_m) \mu_{\psi_l \to X_m}^{\rho_l - 1}(x_m) \prod_{k \in \mathcal{N}_m \setminus \{l\}} \mu_{\psi_k \to X_m}^{\rho_k}(x_m),$$

where $\sum_{\sim x_n}$ denotes summation over all variables except x_n. The beliefs are given by

$$b_n(x_n) \propto \phi_n(x_n) \prod_{l \in \mathcal{N}_n} \mu_{\psi_l \to X_n}^{\rho_l}(x_n). \tag{9.11}$$

Here, the variables ρ_l refer to the appearance probabilities of the factor vertices ψ_l in the collection of trees. A slight abuse of notation has been made, as here $l \in \mathcal{N}_n$ means that there exists a factor ψ_l that has as variable X_n. For URW-BP, all ρ_l are simply set as $\rho_l = \rho$.

For the sake of completeness, message passing rules are also provided when messages from variable vertices to factor vertices are computed. In that case, the message from variable vertex X_n to factor vertex ψ_l, $n \in C_l$ is given by

$$m_{X_n \to \psi_l}(x_n) = \phi_n(x_n) m_{\psi_l \to X_n}^{\rho_l - 1}(x_n) \prod_{k \in \mathcal{N}_n \setminus \{l\}} m_{\psi_k \to X_n}^{\rho_k}(x_n), \tag{9.12}$$

while the message from factor vertex ψ_l to variable vertex X_n, $n \in C_l$ is now

$$m_{\psi_l \to X_n}(x_n) = \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \prod_{m \in C_l \setminus \{n\}} m_{X_m \to \psi_l}(x_m). \tag{9.13}$$

Finally, the belief of variable x_n is given by (9.11), where $\mu_{(\cdot)}(\cdot)$ should be replaced by $m_{(\cdot)}(\cdot)$. These message passing rules can be seen as extension of the sum-product algorithm from [62].

9.2.4 *Discussion of URW-BP*

9.2.4.1 *URW-BP as a Multi-Objective Optimization Solution*

The URW-BP objective function can be written for general interactions as

$$\chi(b) + \sum_{n=1}^{N} \mathcal{H}(b_n) - \rho \sum_{l=1}^{L} \mathcal{I}(b_{C_l}),$$

which can be interpreted as a multi-objective problem with trade-off parameter $\rho \geqslant 0$. The function $\chi(b)$ is linear in $b(\mathbf{x})$, while $\sum_{n=1}^{N} \mathcal{H}(b_n)$ is concave in $b(\mathbf{x})$. Suppose to have two candidate solutions: $b(\mathbf{x})$ and $\tilde{b}(\mathbf{x})$, for

which $\chi(b) = \chi(\tilde{b})$. When $\rho = 1$, the solution with the maximal Bethe entropy will be chosen. Maximizing Bethe entropy corresponds to maximizing the entropies of $b_n(x_n)$ and at the same time minimizing the dependence among variables. When $\rho = 0$, the solution with the maximal entropy of the individual beliefs will be chosen, irrespective of the dependence (mutual information) among the variables. Hence, *by decreasing* ρ, *beliefs are forced to be less concentrated* (i.e., less over-confident).

9.2.4.2 *URW-BP as Optimized TRW-BP*

There exist sufficient, though not necessary conditions, under which URW-BP and optimized TRW-BP coincide. In the following, set cardinality is denoted by $|\cdot|$.

Definition 1 (Symmetric factorization)**.** *Given a factor graph* G*, corresponding to a factorization*

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^{N} \phi_n(x_n|\theta_n) \prod_{l=1}^{L} \psi_l(x_{C_l}|\theta_{C_l}) \tag{9.14}$$

with L *non-trivial*[3] *factors, where* θ_n *and* θ_{C_l} *represents parameters that fully determine the corresponding functions. A factorization is called symmetric when (i)* $|C_l| = |C_k|, \forall k, l \in \{1, \ldots, L\}$ *and* $|\mathcal{N}_n| = |\mathcal{N}_m|, \forall m, n \in \{1, \ldots, N\}$*; (ii)* $\theta_n = \theta_m, \forall m, n \in \{1, \ldots, N\}$ *and* $\theta_{C_l} = \theta_{C_k}, \forall k, l \in \{1, \ldots, L\}$.

Proposition 1. *The optimal* EAPs *in* TRW-BP *for factorizations which are either symmetric, or for which the corresponding factor graph is a tree, are uniform, i.e.,* $\boldsymbol{\rho} = \rho\mathbf{1}$*, for some* $\rho \in [0, 1]$.

The proof follows immediately from symmetry considerations. As a special case, [143, page 2327] points out that for symmetric factorizations with only pairwise interactions, the optimal EAP in TRW-BP should be uniform, and is given by $\rho = (N-1)/L$, where $L = |\mathcal{E}|$. Introducing n_D as the number of pairwise interactions in which each variable is involved,[4] a simple counting argument yields $n_D \approx 2L/N$, so that

$$\rho \approx \min\left(1, \frac{2}{n_D}\left(1 - \frac{1}{N}\right)\right) \tag{9.15}$$

$$\approx \min\left(1, 2/n_D\right). \tag{9.16}$$

Note that for trees, $L = N - 1$, so that $n_D \approx 2$, with corresponding $\rho = 1$.

While symmetric factorizations are not very interesting from a practical perspective, it can be expected that URW-BP should perform close to optimized TRW-BP as long as the factorization exhibits sufficient symmetry or

3 In the sense that $|C_l| \geqslant 2$.

4 Note that n_D is also the node degree in the corresponding Markov random field, and that $n_D + 1$ is the variable vertex degree in the corresponding factor graph.

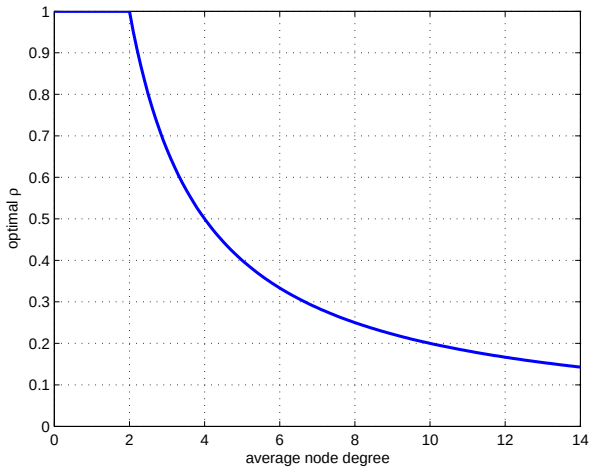


Figure 9.1: Value of ρ as a function of the average node degree according to (9.16).

when the factor graph is sufficiently tree-like. The conditions for sufficient symmetry are not explored here, and are a topic for future research. However, as shown in the following section, the class of factorizations for which URW-BP outperforms BP turns out to be quite large.

In the context of distributed inference, where there are N devices, with a total of L bi-directional communication links among devices, n_D corresponds to the average connectivity, and can be estimated by running a simple consensus algorithm over a network [46]. The relationship (9.16) is shown in Fig. 9.1. It is clear that for high network connectivity, URW-BP is expected to outperform BP.

9.3 CASE STUDIES

In this section, URW-BP will be applied to three practical applications, involving discrete and continuous variables, centralized and distributed processing, and pairwise as well as higher-order interactions. The first application involves a cognitive radio network, where devices determine whether or not the spectrum is being utilized. The second application involves determining positions of wireless devices through cooperation. The goal of every device is to determine its own position based on distance estimates with respect to neighbors and reference nodes. The final application involves decoding of LDPC codes. In all three cases URW-BP is shown to outperform BP.

9.3.1 *Case Study 1: Signal Detection for Cognitive Radio*

9.3.1.1 *Problem Formulation*

In cognitive radio networks, secondary users (SU) are granted access to the spectrum of primary users (PU), when said spectrum is not used by the PU. This allows spatial reuse of scarce spectrum. Prior to using the channel, the SU must sense the spectrum and decide whether or not any PUs are present. As PUs correspond to signal sources with a precise physical location, it is reasonable to assume that two SU that are close to each other are likely to observe the same PU signal state, whereas observations of SU that are far apart are most likely independent. Based on [66], these properties can be captured through (i) binary variables ($x_n \in \{0, 1\}$, where $x_n = 1$ means that the channel is occupied near the location of the n-th SU), (ii) observations $\mathbf{y}_n$ (reflecting the signals collected by the n-th SU) with corresponding likelihoods $\phi_n(x_n) = p(\mathbf{y}_n|x_n)$, and (iii) spatial correlations[5] between x_n and x_m, $\psi_{m,n}(x_n, x_m) = \exp(\lambda_{m,n}\mathbb{I}\{x_n = x_m\})$, where $\lambda_{m,n} \geqslant 0$ generally decreases with the distance between the two SUs. The set of neighbors $\mathcal{N}_n$ in this case is defined implicitly as the set of SU that are within communication range with the n-th SU [109]. The joint a posteriori distribution $p(\mathbf{x}|\mathbf{y})$ is of the form (5.8):

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^{N} p(\mathbf{y}_n|x_n) \prod_{(m,n)\in\mathcal{E}} \exp(\lambda_{n,m}\mathbb{I}\{x_n = x_m\}),$$

where the product over couples (m, n) is only over $m < n$, for which $m \in \mathcal{N}_n$. The goal of the n-th SU is to determine $b_n(x_n) \approx p(x_n|\mathbf{y})$, where $\mathbf{y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$.

The likelihood functions $\phi_n(x_n) = p(\mathbf{y}_n|x_n)$ account for the detection performance of each node n. Assume that each node uses as test statistic for detection a vector of S i.i.d. signal received samples, and that noise and PU signal are both complex Gaussian random variables with variance σ_v^2 and σ_s^2 respectively. Then, it can be readily shown that for $x_n = 0$ (i.e., when no PU is accessing the channel around the n-th SU), $\mathbf{y}_n \sim \mathcal{CN}(0, \sigma_v^2\mathbf{I}_S)$, while for $x_n = 1$ (i.e., when a PU is active), $\mathbf{y}_n \sim \mathcal{CN}(0, (\sigma_v^2 + \sigma_s^2)\mathbf{I}_S)$, where $\mathbf{I}_S$ is the $S \times S$ identity matrix.

Note that distributed signal detection can be seen as a special case of the more general problem of distributed parameter estimation, or multiple hypothesis testing, studied in the context of wireless sensor networks (see, for example, [27, 25, 140]). The proposed method can be extended accordingly to these applications.

5 In general, higher-order interactions (e.g., between x_n, x_m, and x_p) are possible as well, though, similar to [66], these interactions are neglected for simplicity.

9.3.1.2 *Implementation Aspects*

As every SU in the network corresponds to a variable x_n, URW-BP can be implemented distributedly by mapping the factor graph to the communication graph. Thus, the message $\mu_{\psi_{m,n}\to x_m}(x_m)$ from (9.3) in the factor graph is transmitted from the n-th SU to the m-th SU. For binary variables, it is convenient to represent messages with log-likelihood ratios (LLRs). Introducing $\gamma_n = \log(\phi_n(0)/\phi_n(1))$, and accounting for the fact that $\psi_{m,n}^{1/\rho}(0,1) = \psi_{m,n}^{1/\rho}(1,0) = 1$ and $\psi_{m,n}^{1/\rho}(0,0) = \psi_{m,n}^{1/\rho}(1,1) = \exp(\lambda_{m,n}/\rho)$, $\mu_{\psi_{m,n}\to x_m}(x_m)$ can be expressed as a scalar quantity:

$$\begin{aligned} M_{n\to m} &\triangleq \log \frac{\mu_{\psi_{m,n}\to x_m}(0)}{\mu_{\psi_{m,n}\to x_m}(1)} \\ &= \mathbb{U}\left(\frac{\lambda_{m,n}}{\rho}, \gamma_n + (\rho-1)M_{m\to n} + \rho \sum_{k\in\mathcal{N}_n\setminus\{m\}} M_{k\to n}\right), \end{aligned} \tag{9.17}$$

where $\mathbb{U}(a,b) = \log(1+e^{a+b})/(e^a+e^b)$, which is a function that can be computed efficiently for any $a, b \in \mathbb{R}$. Similarly, the beliefs in LLR representation are given by

$$\begin{aligned} B_n &\triangleq \log \frac{b_n(0)}{b_n(1)} \\ &= \log \frac{\phi_n(0)\prod_{m\in\mathcal{N}_n}\mu_{\psi_{m,n}\to x_n}^{\rho}(0)}{\phi_n(1)\prod_{m\in\mathcal{N}_n}\mu_{\psi_{m,n}\to x_n}^{\rho}(1)} \\ &= \gamma_n + \rho\sum_{m\in\mathcal{N}_n} M_{m\to n}. \end{aligned} \tag{9.18}$$

9.3.1.3 *Performance Results*

A metric to evaluate the performance of URW-BP is KLD between belief and true posterior, defined as

$$\mathrm{KLD}_n \triangleq \mathrm{KL}\left(b_n(x_n)\|p(x_n|\boldsymbol{y})\right), \tag{9.19}$$

computed for each node in the network and averaged over all nodes. First a small network of 4 nodes is considered, with correlation factors $\lambda_{n,m}$ modeled as uniformly distributed random variables between 0.2 and 4. Nodes are deployed randomly in a circular area of unit diameter. Then, define R as the communication range of nodes in the network: if $R = 1$, all nodes can communicate with each other, therefore many loops are present in the graph; on the contrary, low values of R result in fewer loops. The first graph in Fig. 9.2 shows the average KLD of beliefs computed through standard BP and URW-BP as a function of ρ for different values of R. It can be observed that URW-BP outperforms standard BP (corresponding to $\rho = 1$), especially for large values of R. Also, the best value of ρ (denoted as ρ^*) tends to decrease as $R \to 1$ as an effect of the increasing number of loops in the graph.

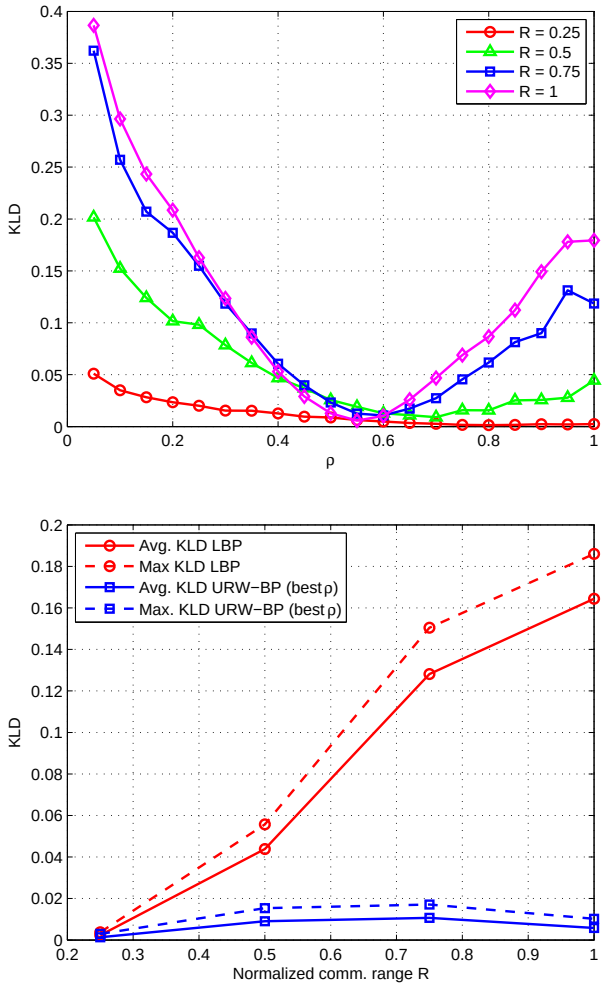


Figure 9.2: Average KLD vs. ρ (top) and vs. normalized communication range (bottom) for a simple network of 4 nodes, $\lambda_{m,n} \sim \mathcal{U}(0.2, 4)$. The value $\rho = 1$ corresponds to standard BP.

In the the second panel of Fig. 9.2, average and maximum KLD are plotted as function of the communication range, for URW-BP and standard BP.

Fig. 9.3 combines the results of simulations performed on networks of different size (N from 4 through 15) and shows the best value of ρ (according to average KLD) as a function of the average degree n_D. Results confirm the dependency of ρ^* on the average degree of the graph, as discussed in Sec. 9.2.4. In particular, the approximate expression (9.16) proves to be increasingly accurate as $n_D \gg 1$.

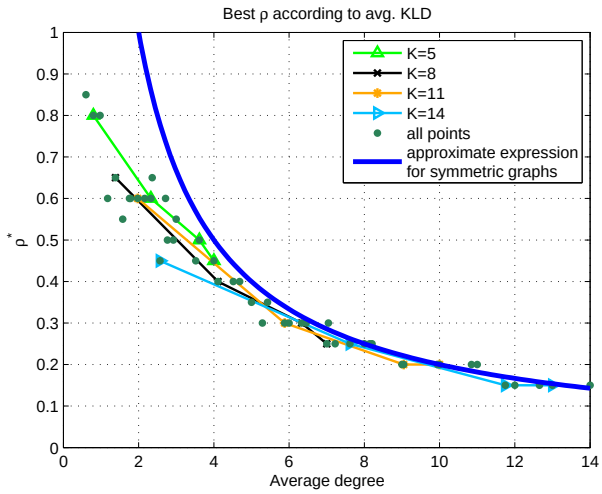


Figure 9.3: Best EAP (ρ^*) based on average KLD vs. average degree n_D.

9.3.2 *Case Study 2: Cooperative Positioning*

9.3.2.1 *Problem Formulation*

The second considered application of URW-BP is cooperative positioning in wireless networks. In this case, the goal of each node (referred to as targets) is to estimate its own position based on a set distance measurements with nearby targets and and few fixed reference nodes (referred to as anchors), assuming a fully distributed architecture. These distance estimates can be obtained using well-known measurement techniques (e.g., time of arrival, received signal strength). Contrary to the previous case, *continuous* variables occur in this problem. Also contrary to the previous case where over-confidence of the beliefs is not detrimental as long as the estimates are correct, here the situation is quite different: for safety-critical applications, e.g., tracking of robots or unmanned aerial vehicles (UAVs), it is important not only to have good position estimates, but also to know the uncertainty. When a belief is overconfident, a target may take improper actions (e.g., crash into another UAV).

The considered scenario consists of N targets and minimum 3 anchors scattered randomly in a planar region, and denote the two-dimensional location of the n-th node (target or anchor) by x_n. Target n obtains a noisy measurement y_{mn} of its distance from node m, given by

$$y_{mn} = \|x_m - x_n\| + v_{mn}, \tag{9.20}$$

where v_{mn} represents the noise drawn from some distribution p_v (e.g., Gaussian or the empirical distribution found from real measurements). As in the previous case study, it is assumed that nodes can only communicate with nearby devices within a predefined radius $R > 0$. Thus, y_{mn} is available to target n only when $\|x_m - x_n\| < R$. Hence, $m \in \mathcal{N}_n$ when $\|x_m - x_n\| < R$.

Letting $\phi_n(x_n) = p_n(x_n)$ be the a priori distribution of the position of the n-th target, $\psi_{m,n}(x_m, x_n) = p(y_{mn}|x_n, x_m)$ the likelihood function, and collecting the target positions in a vector $\mathbf{x}$, and the distance estimates in a vector $\mathbf{y}$, the a posteriori distribution $p(\mathbf{x}|\mathbf{y})$ is of the form (5.8):

$$p(\mathbf{x}|\mathbf{y}) \propto \prod_{n=1}^{N} p_n(x_n) \prod_{(m,n)\in\mathcal{E}} p(y_{mn}|x_n, x_m),$$

where the product over couples (m, n) is only over nodes m within communication range of target n, i.e., $m \in \mathcal{N}_n$. The goal of target n is to determine its belief $b_n(x_n) \approx p(x_n|\mathbf{y})$. From the belief, the target can easily find a point estimate (e.g., the mean of $b_n(x_n)$) and associated uncertainty (e.g., the covariance matrix, or any percentile).

9.3.2.2 *Implementation Aspects*

Since the variables are now continuous, all sums in message update equations in Section 9.2 have to be replaced with integrals. As the localization problem is nonlinear and potentially non-Gaussian, exact message representation and updating is not tractable. Therefore, a non-parametric version of URW-BP (URW-NBP) is adopted, where the beliefs and message update equations are performed using particle-based approximations (see [49, 123, 158] for more details on NBP).

9.3.2.3 *Performance Results*

A network with 4 anchors and $N = 25$ targets is considered, in a square 20 m by 20 m deployment area. For simplicity, it is assumed that the noise v_{mn} has a zero-mean Gaussian distributionwith standard deviation 30 cm. Ten iterations of message passing are performed, representing the messages with 500 particles. Performance is measured in terms of the average ℓ_1 positioning error and the average trace of the covariance of the belief. The ℓ_1 error for target n is given by $\|x_n - \hat{x}_n\|$, where $\hat{x}_n$ is the mean of $b_n(x_n)$. The results are averaged over 100 Monte Carlo runs.

Fig. 9.4 shows the performance of URW-BP as a function of $\rho \in [0.1, 1.5]$ for $R = 6.6$ m and $R = 16$ m after the tenth iteration. It turns out that for low connectivity, the ℓ_1 error is relatively insensitive to ρ for any $\rho > 0.4$. However, the beliefs become more concentrated as ρ increases. Hence, reducing ρ results in a more robust algorithm. When the connectivity is increased ($R = 16$), the best value of $\rho \approx 0.3$, while $\rho = 1$ induces around 20% additional error. Again, more concentrated beliefs are observed with increasing ρ. Overall, the differences between BP and URW-BP are small.

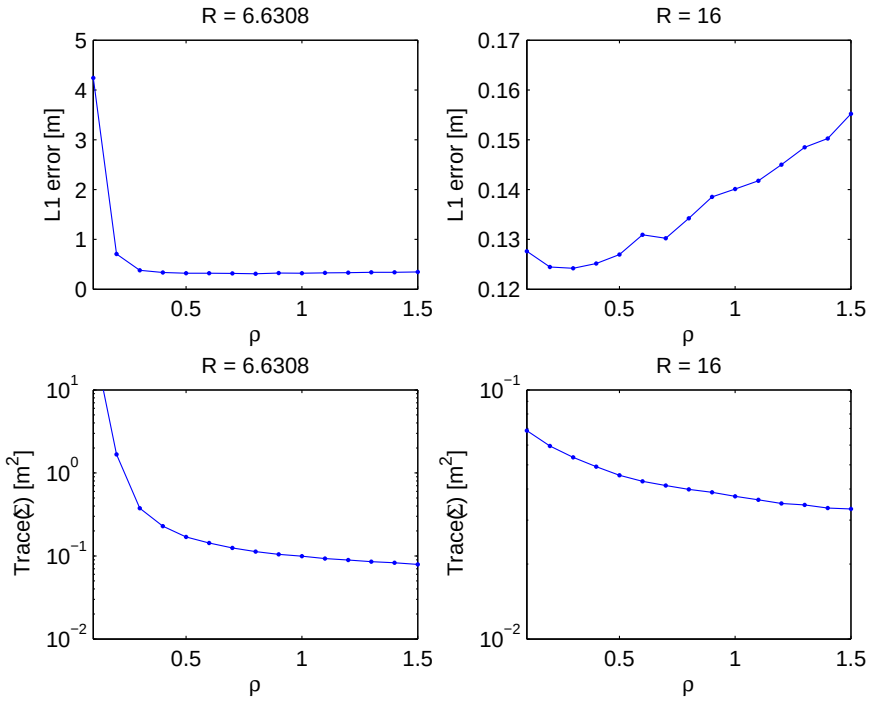


Figure 9.4: Positioning performance as a function of ρ for different communication radii. The first row shows the average ℓ_1 error, and the second row the trace of the covariance matrix (denoted by $\boldsymbol{\Sigma}$) of the belief.

9.3.3 *Case Study 3: LDPC Decoding*

9.3.3.1 *Problem Formulation*

Consider an LDPC code with an $L \times N$ sparse parity check matrix $\mathbf{H}$. Let $\mathbf{x}$ denote the transmitted codeword and $\mathbf{y}$ the observation over a memoryless channel, with known $p(y_n|x_n)$. The a posteriori distribution of interest is now

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= \prod_{n=1}^{N} p(y_n|x_n)\mathbb{I}\{\mathbf{H}\mathbf{x} = \mathbf{0}\} \\ &= \prod_{n=1}^{N} p(y_n|x_n) \prod_{l=1}^{L} \mathbb{I}\left\{\sum_{m\in C_l} x_m = 0\right\}, \end{aligned}$$

where the summation is in the binary field, and C_l is the index set corresponding to the non-zero elements of the l-th row in $\mathbf{H}$. Clearly, the above distribution can be associated with (9.10) through $\phi_n(x_n) \leftrightarrow p(y_n|x_n)$ and $\psi_l(x_{C_l}) \leftrightarrow \mathbb{I}\left\{\sum_{m\in C_l} x_m = 0\right\}$.

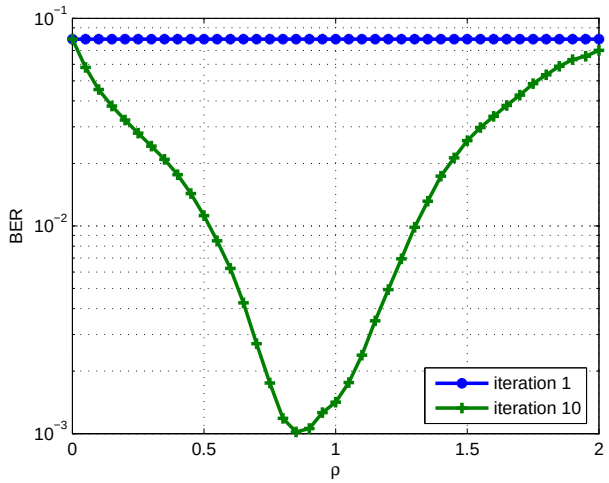


Figure 9.5: Performance of URW-BP as a function of the scalar parameter ρ for an LDPC code, at a fixed SNR ($E_b/N_0 = 3$dB). BP corresponds to $\rho = 1$.

9.3.3.2 *Implementation Aspects*

Due to the nature of $\psi_l(x_{C_l})$, $\psi_l^{1/\rho}(x_{C_l}) = \psi_l(x_{C_l})$, irrespective of ρ. Hence, the messages from check nodes to variable nodes are unchanged with respect to standard BP. Messages can be represented efficiently in the log-domain, similar to the cognitive radio problem. It turns out that the message from variable node X_n to check node ψ_l, expressed as a LLR, is given by (see also (9.12))

$$M_{X_n \to \psi_l} = M_{ch,n} + \rho \sum_{k \in \mathcal{N}(X_n)} M_{\psi_k \to X_n} - M_{\psi_l \to X_n},$$

where $\lambda_{ch,n} = \log p(y_n|x_n = 1)/p(y_n|x_n = 0)$. For additional details see [160].

9.3.3.3 *Performance Results*

Fig. 9.5 shows the bit error rate (BER) performance of a rate $1/2$ LDPC code with $N = 256$ and $L = 128$ at a fixed SNR as a function of ρ, over an AWGN channel. It is observed that $\rho = 1$ does not yield the best performance and that the global minimum in the BER is achieved by $\rho \approx 0.85$. Fig. 9.6 compares the performance of BP and URW-BP (with $\rho = 0.85$) in terms of BER vs. SNR for the same LDPC code. The result is that URW-BP outperforms BP at high SNR. The difference in BER between TRW-BP and URW-BP is not large, but still significant considering that a real LDPC code has been used, designed such that loops are long and have limited impact on BP decoding. The performance gap can be much greater in case of non-optimized graph configurations, i.e., with many short loops.

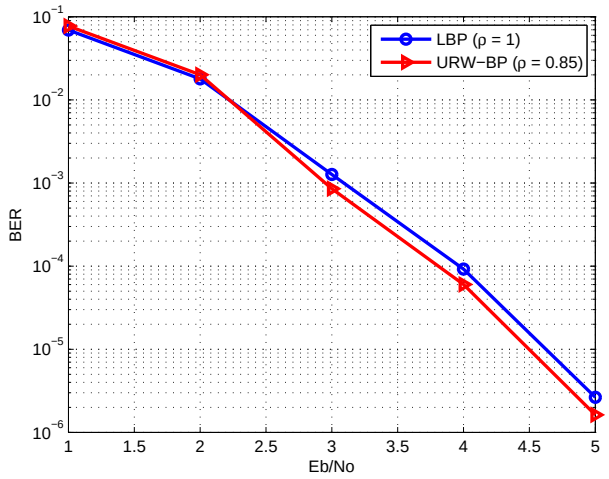


Figure 9.6: Performance of BP and URW-BP (with $\rho = 0.85$) as a function of the SNR for an LDPC code.

Part IV

CONCLUSION

SUMMARY AND CONCLUSIONS

Various examples of application of statistical techniques to wireless communications have been presented in this thesis. The focus has been on *distributed inference problems* in wireless networks, where multiple nodes (sensors) have to estimate the value of one or more hidden variables. Different specific problems – signal detection, cooperative localization, identification of misbehaving users – have been formulated under a common framework.

RMT techniques have been applied in the case of *homogeneous* inference problems, where all sensors observe the same hidden variable. In this case, the received signal covariance matrix (constructed from datasets received at multiple sensors) is a key test statistic to estimate the hidden variable of interest. RMT provides mathematical tools to analyze the statistical properties of such matrix and, in particular, of its eigenvalues.

Such a methodology has been useful to characterize the performance (namely, false-alarm and detection probability) of several eigenvalue-based detectors. Results based on RMT, and in particular on asymptotical eigenvalue distributions, have been proven to be more accurate than previously proposed approaches, based on point-wise (almost sure) convergence of eigenvalues. Expressions of false-alarm probability have been then used to determine the decision threshold as a function of the target false-alarm rate. The main application of these results is multi-sensor spectrum sensing in cognitive radio networks, where reliable detection of weak signals is needed to enable interference-free coexistence between primary and secondary users in the same frequency bands.

As a case-study, results on signal detection have been applied to low-power networks (e.g., WSNs) coexisting with other transmitters in unlicensed bands. Both theoretical and experimental results have been presented, showing that a substantial performance increase (in terms of throughput and packet loss rate) can be achieved by adopting approaches based on the CR principle. Also, specific expressions of signal detection performance have been derived for the considered scenario, taking into account discontinuous channel occupation by PU transmitters.

Probabilistic graphical models (FGs and MRFs) have been adopted, on the other hand, to address *heterogeneous* inference problems, where a different hidden variable exists for each node in the network. In this case, the statistical problem structure can be mapped to a graph, and hidden variables can be estimated by iterative algorithms such as BP (SPA).

If nodes in the network are collaborative and statistical interactions among variables are pairwise, then it is possible to implement the SPA or other

message passing algorithms in a *decentralized* fashion, such that messages defined over the edges of a FG or MRF are exchanged by network nodes as physical packets over the wireless channel. Such techniques, referred to as *network message passing* algorithms, provide increased efficiency, scalability, and robustness compared to centralized schemes. Two examples of network message passing algorithms have been considered in this thesis: *hybrid (GNSS+terrestrial) cooperative localization* and *distributed signal detection in heterogeneous environments*. In both cases, BP-based solutions have been shown to provide substantial performance improvements over traditional methods.

Examples of inference problems solved by BP, but in a centralized setting, have been considered as well. An example is cooperative signal detection in the presence of malicious sensor nodes. This problem can be formulated as an inference problem involving at the same time a *homogeneous* hidden variable (signal presence, assumed as common to all nodes) and multiple *heterogeneous* variables, modeling the behavior of possibly malicious users. The problem is then mapped to a FG, where all hidden variables are jointly estimated by the fusion center via iterative BP over the FG. As a result, it has been shown that the sensing error probability, after a sufficient number of time slots, is close to that of an ideal MAP estimator with exact knowledge of the attack model.

Finally, motivated by the use of BP in several distributed inference problems, a variation of the BP algorithm has been introduced, where messages exchanged among nodes are weighted by an exponent depending on the average graph degree. The new algorithm, named URW-BP, exhibits superior performance compared to standard BP when the graph contains cycles. URW-BP has been applied to signal detection, localization, and LDPC decoding. A significant performance improvement compared to traditional BP has been observed especially in the case of signal detection. In addition, URW-BP can be implemented as a network message passing algorithm (decentralized implementation), similar to BP.

BIBLIOGRAPHY

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, S. Mohanty, "Next Generation / Dynamic Spectrum Access / Cognitive Radio Wireless Networks: A Survey", *ELSEVIER International Journal of Computer and Telecommunications Networking (ComNet)*, no. 50, May 2006, pp. 2127-2159.

[2] T.W. Anderson, *An introduction to multivariate statistical analysis*, 3rd. ed., New York, Willey, 2003.

[3] Z. D. Bai, "Methodologies in spectral analysis of large-dimensional random matrices, a review", *Statistica Sinica*, vol.9, pp.611-677, 1999.

[4] Z. Bai, J. Yao, J, "Central limit theorems for eigenvalues in a spiked population model", *Ann. Inst. H. Poincaré'*, vol. 44, no. 3, pp. 447–474, 2008.

[5] J. Baik, G. Ben Arous, G. and S. Péché, "Phase transition of the largest eigenvalue for non-null complex sample covariance matrices", *Ann. Probab.*, vol. 33, no. 5, pp. 1643-1697, 2005.

[6] J. Baik and J. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models", *Journ. of Mult. Anal.*, vol. 97, pp. 1382-1408, 2006.

[7] O. Besson, L.L. Scharf, "CFAR matched direction detector", *IEEE Trans. on Sig. Proc.*, vol.54, no.7, pp.2840-2844, July 2006.

[8] R. S. Blum, S. A. Kassam, H. V. Poor, "Distributed Detection with Multiple Sensors: Part II – Advanced Topics", *Proc. of the IEEE*, vol. 85, no. 1, Jan. 1997.

[9] P. Bianchi, J. Najim, G. Alfano, and M. Debbah, "Asymptotics of eigenbased collaborative sensing", *Proc. IEEE Information Theory Workshop (ITW 2009)*, Taormina, Italy, Oct. 2009.

[10] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory", http://arxiv.org/abs/0910.0827.

[11] J. Broch, D.A. Maltz, D.B. Johnson, Y.C. Hu, J. Jetcheva J, "A performance comparison of multi-hop wireless ad hoc network routing protocols", *Proc. 4th ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom98)*, October 1998.

[12] M. Brookes, "The Matrix Reference Manual", online: http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html, 2005.

[13] D. Cabric, S.M. Mishra, R.W. Brodersen, "Implementation Issues in Spectrum Sensing for Cognitive Radios", *Proc. IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2004.

[14] M. Caceres, F. Penna, H. Wymeersch, R. Garello, "Hybrid GNSS-Terrestrial Positioning via Distributed Belief Propagation", *IEEE GLOBECOM 2010*, Miami, FL, USA, Dec. 6-10, 2010

[15] M.A. Caceres, F. Sottile, R. Garello and M.A. Spirito "Hybrid GNSS-ToA Localization and Tracking via Cooperative Unscented Kalman Filter" in *Proceedings of the IEEE 21st International Conference on Personal, Indoor and Mobile Radio Communications (PIMRC) Workshops*, Istanbul, Turkey, September 26–29 2010, pp. 271–275.

[16] M. Caceres, F. Penna, H. Wymeersch, R. Garello, "Hybrid Cooperative Positioning based on Distributed Belief Propagation", accepted for publication in *IEEE Journal on Selected Areas of Communications*, to appear (Dec. 2011).

[17] L.S Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative spectrum sensing using random matrix theory", *Proc. 3rd International Symposium on Wireless Pervasive Computing (ISWPC) 2008*, pp.334-338, May 2008.

[18] J.-C. Chen, C.-S. Maa, Y.-C. Wang, and J.-T. Chen, "Mobile Position Location Using Factor Graphs," *IEEE Communications Letters*, vol. 7, pp. 431–433, September 2003.

[19] J. Chen, Y. Wang, C. Maa, and J. Chen, "Network-side Mobile Position Location Using Factor Graphs," in *IEEE Transactions on Wireless Communications*, vol. 5, 2006, pp. 2696–2704.

[20] H. S. Chen, W. Gao, and D. G. Daut, "Signature based spectrum sensing algorithms for IEEE 802.22 WRAN," *IEEE International Conference on Communications (ICC 07)*, pp. 6487-6492, Glasgow, Scotland, UK, June 2007.

[21] P. Chen, M.C. Wicks and R.S. Adve, "Development of a statistical procedure for detecting the number of signals in a radar measurement", *IEE Proc. on Radar Sonar Navigation*, vol 148, no. 4, pp. 219-226, 2001.

[22] M. Chiang, "Distributed network control through sum product algorithm on graphs", *Proc. of IEEE Global Telecommunications Conference (Globecom)*, vol. 3, pp. 2395-2399, 2003.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, second ed., 2006.

[24] J. H. Curtiss, "On the Distribution of the Quotient of Two Chance Variables", *The Annals of Mathematical Statistics*, vol. 12, no. 4, pp. 409-421, 1941.

[25] V. Delouille, R. Neelamani, V. Chandrasekaran, and R. G. Baraniuk, "The embedded triangles algorithm for distributed estimation in sensor networks", in *Proc. 2003 IEEE Workshop Stat. Signal Process.*, 2003.

[26] M. Dieng, *RMlab ver. 0.02*: http://math.arizona.edu/~momar/, 2006

[27] A. Dogandzic, B. Zhang, "Distributed Estimation and Detection for Sensor Networks Using Hidden Markov Random Field Models", *IEEE Trans. on Signal Processing*, vol. 54, no. 8, Aug. 2006.

[28] A.W. Eckford, "Channel estimation in block fading channels using the factor graph EM algorithm", in *Proc. 22nd Biennial Symposium on Communications*, Kingston, Canada, 2004.

[29] N. El Karoui, "A rate of convergence result for the largest eigenvalue of complex white Wishart matrices", *Ann. Prob.*, vol. 36, no. 6, pp. 2077- 2117, 2006.

[30] S. Enserink, D. Cochran, "A cyclostationary feature detector," *28th Asilomar Conference on Signals, Systems and Computers*, pp.806-810, Oct. 1994.

[31] E.B. Ermis, V. Saligrama, "Distributed Detection in Sensor Networks With Limited Range Multimodal Sensors", *IEEE Trans. on Signal Processing*, vol. 58, no. 2, Feb. 2010.

[32] C. Falsi, D. Dardari, L. Mucchi, and M. Win, "Time of arrival estimation for UWB localizers in realistic environments," in *EURASIP Journal of Applied Signal Processing*, 2006:Article ID 32 082, 1–13.

[33] FCC Second Memorandum of Order, Sept. 23, 2010: http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-10-174A1.pdf

[34] O. N. Feldheim, S. Sodin, "A universality result for the smallest eigenvalues of certain sample covariance matrices", preprint: http://arxiv.org/abs/0812.1961, 2008.

[35] D. Féral, S. Péché, "The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case", preprint: http://arxiv.org/abs/0812.2320, 2008.

[36] E. Fishler, M. Grosmann, H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis", *IEEE Transactions on Signal Processing*, vol.50, no.5, pp.1027-1036, May 2002.

[37] B. J. Frey and N. Jojic, "A comparison of algorithms for inference and learning in probabilistic graphical models", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1392-1416, Sept. 2005.

[38] M. Gandetto, C. Regazzoni, "Spectrum Sensing: A Distributed Approach for Cognitive Terminals", *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, April 2007.

[39] A. Ghasemi, E. S. Sousa, "Spectrum Sensing in Cognitive Radio Networks: Requirements, Challenges and Design Trade-offs", *IEEE Communications Magazine*, April 2008, pp.32-39.

[40] A. Ghasemi, E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments", *Proc. of IEEE DySPAN*, 2005.

[41] J. M. Hasik and Michael R. Rip. "GPS at war: a ten-year retrospective", in *Proceedings of the 14th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS)*, Salt Lake City, UT, September 11–14, 2001, pp. 2406–2417.

[42] H. Hashemi, "The indoor radio propagation channel," in *Proceedings of the IEEE*, vol. 81 pp. 943–968, July 1993.

[43] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications", *IEEE Trans. Commun.*, vol. 23, no. 2, pp. 201–220, 2005.

[44] G. Heinrichs, P. Mulassano, and F. Dovis, "A hybrid positioning algorithm for cellular radio networks by using a common rake receiver," in *Symp. on Personal, Indoor and Mobile Radio*, 2004, pp. 2347–2351.

[45] V. Hinkley. "On the Ratio of Two Correlated Normal Random Variables". *Biometrika*, vol. 56, no. 3, pp. 635-639, Dec. 1969.

[46] Y.-W. Hong, A. Scaglione, and P. K. Varshney, "A communication architecture for reaching consensus in decision for a large network", *IEEE Statistical Signal Processing Workshop (SSP)*, Paris, France, pp. 1220-1225, July 2005.

[47] IEEE 802.22, "Draft Standard for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control(MAC) and Physical Layer(PHY) Specifications: Policies and Procedures for Operation in the TV Bands", *IEEE P802.22-D1.1*, July 2008.

[48] A. T. Ihler, J. Fisher III, R. Moses, and A. Willsky, "Nonparametric belief propagation for self-calibration in sensor networks," in *Proc. of the 3rd Intl. Symposium on Information Processing in Sensor Networks*, 2004.

[49] A. T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks", *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 23, pp. 809-819, Apr. 2005.

[50] A. T. Ihler, "Inference in Sensor Networks: Graphical Models and Particle Methods," PhD Thesis, MIT, 2005.

[51] K. Johansson, "Shape fluctuations and random matrices", *Comm. Math. Phys.*, vol. 209, no. 12, pp. 437–474, 2000.

[52] D. E. Johnson, F. A. Graybill, "An analysis of a two-way model with interaction and no replication", *J. Amer. Stat. Assoc.* no. 67, pp. 862-868, 1972.

[53] I. Johnstone, "On the distribution of the largest eigenvalue in principal component analysis", *Ann. Statist.*, pp. 295-327, 2001.

[54] E. Kaplan and C. Hegarty. *Understanding GPS: Principles and Applications.* Artech House Mobile Comm. Series, 2006.

[55] H. Khaleel, C. Pastrone, F. Penna, M. A. Spirito, R. Garello, "Impact of Wi-Fi Traffic on the IEEE 802.15.4 Channels Occupation in Indoor Environments", *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Turin, Italy, Sept. 14-18, 2009

[56] H. Khaleel, F. Penna, C. Pastrone, R. Tomasi, M. Spirito, "Distributed Spectrum Sensing and Channel Selection in Opportunistic Wireless Personal Area Networks", *ACM MobiOpp 2010 - Demo Session*, Pisa, Italy, Feb. 22-24, 2010 - *Best Demo Award*

[57] H. Khaleel, F. Penna, C. Pastrone, R. Tomasi, M. A. Spirito, "Frequency-Agile Wireless Sensor Networks: Design and Implementation", accepted for publication in *IEEE Sensors Journal*, 2011.

[58] A. Kortun, T. Ratnarajah, M. Sellathurai, and C. Zhong, "On the performance of eigenvalue-based spectrum sensing for cognitive radio", *Proc. IEEE DySPAN*, Singapore, Apr. 2010.

[59] H. Krim, M. Viberg, "Two Decades of Array Signal Processing Research: The Parametric Approach", *IEEE Signal Processing Magazine*, vol. 24, no. 4, July 2006.

[60] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data", *Chemometrics and Intelligent Laboratory Systems*, vol. 94, pp. 19-32, 2008.

[61] S. Kritchman, B. Nadler, "Non-Parametric Detections of the Number of Signals: Hypothesis Testing and Random Matrix Theory", *IEEE Trans. on Signal Processing*, vol. 57, no. 10, pp. 3930–3941, 2009.

[62] F.R. Kschischang, B.J. Frey, H. A. Loeliger, "Factor Graphs and the Sum-Product Algorithm," in *IEEE Transactions on Information Theory*, vol. 47, no. 2, Feb. 2001.

[63] K. Kuusniemi, G. Lachapelle, and J. Takala, "Position and Velocity Reliability Testing in Degraded GPS Signal Environments," *GPS Solutions*, pp. 226–237, October 2004.

[64] Lawley, "Tests of significance for the latent roots of covariance and correlation matrices", *Biometrika*, vol. 43, pp. 128-136, 1956.

[65] J. Lei, R. Yates, P. Spasojevic, and L. Greenstein, "Cooperative sensing of primary users in cognitive radio networks based on message passing", in *43rd Annual Conference on Information Sciences and Systems (CISS)*, pp. 568-573, 2009.

[66] H. Li, "Cooperative Spectrum Sensing via Belief Propagation in Spectrum-Heterogeneous Cognitive Radio Systems", *IEEE Wireless Communications and Networking Conference (WCNC)*, Sydney, Australia, April 2010.

[67] Y.-C. Liang, Y. Zeng, E. C.Y. Peh, A. T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks", *IEEE Trans. on Wireless Communications*, Vol.7, No. 4, 2008.

[68] J. Lien, J. Ferner, W. Srichavengsup, H. Wymeersch, M. Win, "Practical Cooperative Localization and Tracking," submitted to *IEEE Transactions on Wireless Communications*, 2009. Available at http://dl.dropbox.com/u/14551660/LieFerSriWymWinv11_jl.pdf

[69] H. Q. Liu, F. K. W. Chan, and H. C. So, "Non-Line-of-Sight Mobile Positioning Using Factor Graphs," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2009.

[70] H. A. Loeliger, "An introduction to factor graphs", *IEEE Signal Processing Magazine*, vol. 21, 2004.

[71] J. Ma, G. Zhao, Y. Li, "Soft Combination and Detection for Cooperative Spectrum Sensing in Cognitive Radio Networks", *IEEE Transactions on Wireless Communications*, vol.7, no.11, pp.4502-4507, Nov. 2008

[72] Z. Ma, "Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Wishart matrices", *Bernoulli*, to appear.

[73] M. Maida, J. Najim, P. Bianchi, and M. Debbah, "Performance Analysis of Some Eigen-based Hypothesis Tests for Collaborative Sensing", *Proc. IEEE Workshop on Statistical Signal Processing*, Cardiff, UK, 2009.

[74] Y. Mao, F. Kschischang, B. Li, and S. Pasupathy, "A factor graph approach to link loss monitoring in wireless sensor networks", *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 23, no. 4, pp. 820-829, 2005.

[75] V. A. Marchenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices", *Math USSR-Sbornik*, vol.1, pp.457-483, 1967

[76] R. J. McEliece, D.J.C. MacKay, J.-F. Cheng, "Turbo Decoding as an Instance of Pearl's Belief Propagation Algorithm", *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 16, no. 2, Feb. 1998.

[77] MEMSIC TelosB datasheet, http://www.memsic.com/support/documentation/wireless-sensor-networks/category/7-datasheets.html?download=152

[78] C. Mensing and S. Plass, "Positioning Based on Factor Graphs," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–12, September 2007.

[79] C. Mensing, S. Sand, and A. Dammann, "GNSS Positioning in Critical Scenarios: Hybrid Data Fusion with Communications Signals," in *Proceedings of the International Workshop on Synergies in Communications and Localization*, Dresden, Germany, 2009.

[80] S. Mishra, A. Sahai, R. Brodersen, "Cooperative sensing among cognitive radios", *Proc. of IEEE ICC*, 2006.

[81] J. Mitola, G. Q. Maguire, "Cognitive Radios: Making Software Radios More Personal", *IEEE Personal Commun.*, vol. 6, no. 4, pp. 13-18, 1999.

[82] J.M. Mooij, H.J. Kappen, "Sufficient Conditions for Convergence of the Sum-Product Algorithm", *IEEE Trans. on Inform. Theory*, vol. 53, no. 12, Dec. 2007.

[83] G. Moustakides, "Finite Sample Size Optimality of GLR Tests", *submitted to IEEE Trans. on Information Theory*, 2009. Preprint: http://arxiv.org/abs/0903.3795.

[84] R. J. Muirhead, "Latent roots and matrix variates: A review of some asymptotic results," *Ann. Stat.*, vol. 6, no. 1, pp. 5-33, 1978.

[85] K. Murphy, Y. Weiss, and M. Jordan, "Loopy belief propogation for approximate inference: an empirical study", *Conf. on Uncertainty in Artificial Intelligence*, 1999.

[86] B. Nadler, "Nonparametric Detection of Signals by Information Theoretic Criteria: Performance Analysis and an Improved Estimator", *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2746-2756, May 2010.

[87] B. Nadler, "On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart Matrix", *J. of Mult. Anal.*, vol. 102, pp. 363–371, 2010.

[88] B. Nadler, F. Penna, R. Garello, "Performance of Eigenvalue-based Signal Detectors with Known and Unknown Noise Level", *IEEE ICC*, Kyoto, Japan, June 2011.

[89] J. Neyman, E. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Philosophical Transactions of the Royal Society of London*, Series A, vol. 231: pp. 289-337, 1933.

[90] B. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations", *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5491-5499, 2008.

[91] H. Niu, M. Shen, J. A. Ritcey, and H. Liu, "A factor graph approach to iterative channel estimation and LDPC decoding over fading channels", *IEEE Trans. on Wireless Communications*, vol. 4, pp. 1345-1350, July 2005.

[92] R. Olfati-Saber, R.M. Murray, "Consensus problems in networks of agents with switching topology and time-delays", *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520-1533, Sept. 2004.

[93] D. Pados, P. Papantoni-Kazakos, D. Kazakos, A. Koyiantis, "On-line Threshold Learning for Neyman-Pearson Distributed Detection", *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 24, no. 10, Oct. 1994.

[94] M. Paskin, C. Guestrin, and J. McFadden, "A robust architecture for distributed inference in sensor networks", in *Fourth International Symposium on Information Processing in Sensor Networks*, pp. 55-62, 2005.

[95] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model", *Statistica Sinica,* 17:1617–1642, 2007.

[96] S. Péché, "Universality results for largest eigenvalues of some sample covariance matrix ensembles", preprint: http://arxiv.org/abs/0705.1701, 2007.

[97] F. Penna, C. Pastrone, H. Khaleel, M. A. Spirito, R. Garello, "Characterization of Wi-Fi interference for dynamic channel allocation in WPANs", *Reliable Communications for Short-Range Wireless Systems* (edited by I. Guvenc et al.), Cambridge University Press, 2011.

[98] F. Penna, R. Garello, "Decentralized Neyman-Pearson Test with Belief Propagation for Peer-to-peer Collaborative Spectrum Sensing", submitted to *IEEE Transactions on Wireless Communications*, 2011.

[99] F. Penna, Y. Sun, L. Dolecek, D. Cabric, "Detecting and Counteracting Statistical Attacks in Cooperative Spectrum Sensing", submitted to *IEEE Transactions on Signal Processing* (under second round of review), 2011.

[100] F. Penna, R. Garello, "Detection of Discontinuous Signals for Cognitive Radio Applications", *IET Communications*, vol. 5, no. 10, pp. 1453–1561, Aug. 2011.

[101] F. Penna, M. A. Caceres, H. Wymeersch, "Cramér-Rao Bound for Hybrid GNSS-Terrestrial Cooperative Positioning", *IEEE Communications Letters*, vol. 14, no.11, pp. 1005–1007, November 2010.

[102] F. Penna, R. Garello, M. A. Spirito, "Cooperative Spectrum Sensing based on the Limiting Eigenvalue Ratio Distribution in Wishart Matrices", *IEEE Communications Letters*, vol.13, no.7, pp.507–509, July 2009.

[103] F. Penna, D. Cabric, "Bounds and Tradeoffs for Cooperative DoA-only Localization of Primary Users", to be presented at *IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011.

[104] F. Penna, Y. Sun, L. Dolecek, D. Cabric, "Joint Spectrum Sensing and Detection of Malicious Nodes via Belief Propagation", to be presented at *IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011.

[105] F. Penna, J. Wang, D. Cabric, "Cooperative localization of primary users by directional antennas or antenna arrays: Challenges and design issues", *IEEE International Symposium on Antennas and Propagation (APS-URSI)*, Spokane, WA, USA, July 2011.

[106] F. Penna, H. Wymeersch, V. Savic, "Uniformly Reweighted Belief Propagation for Distributed Bayesian Hypothesis Testing", *IEEE Statistical Signal Processing Workshop (SSP)*, Nice, France, June 2011.

[107] F. Penna, R. Garello, M. A. Spirito, "Likelihood-Ratio Propagation and Consensus in Wireless Networks with Markov Random Field Models" *IEEE GLOBECOM - Workshop on Mobile Computing and Emerging Communication Networks*, Miami, FL, USA, Dec. 6-10, 2010

[108] F. Penna, R. Tomasi, H. Khaleel, C. Pastrone, M. Spirito, "On Spectrum Sensing Duration in Cognitive Wireless Sensor Networks", *3rd International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART)* (invited), Rome, Italy, Nov. 8-10, 2010

[109] F. Penna, R. Garello, M. A. Spirito, "Distributed Inference of Channel Occupation Probabilities in Cognitive Networks via Message Passing", *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Singapore, April 6-9, 2010

[110] F. Penna, R. Garello, M. A. Spirito, "Probability of Missed Detection in Eigenvalue Ratio Spectrum Sensing", *5th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Marrakech, Morocco, Oct. 12-14, 2009

[111] F. Penna, R. Garello, D. Figlioli, M. A. Spirito, "Exact Non-asymptotic Threshold for Eigenvalue-based Spectrum Sensing", *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, Hannover, Germany, Jun. 22-24, 2009

[112] F. Penna, C. Pastrone, M. A. Spirito, R. Garello, "Energy Detection Spectrum Sensing with Discontinuous Primary User Signal", *IEEE International Conference on Communications (ICC)*, Dresden, Germany, Jun. 14-18, 2009

[113] F. Penna, C. Pastrone, M. A. Spirito, R. Garello, "Measurement-based Analysis of Spectrum Sensing in Adaptive WSNs under Wi-Fi and Bluetooth Interference", *IEEE 69th Vehicular Technology Conference (VTC)*, Barcelona, Spain, Apr. 26-29, 2009

[114] F. Penna, C. Pastrone, M. A. Spirito, R. Garello, "An Experimental Study on Spectrum Sensing for Cognitive WSNs under Wi-Fi Interference", *21st Wireless World Research Forum (WWRF) meeting*, Stockholm, Sweden, Oct. 13-15, 2008

[115] M. D. Penrose, *Random Geometric Graphs*, Oxford University Press, 2003.

[116] G. Retscher and Q. Fu, "Integration of RFID, GNSS and DR for Ubiquitous Positioning in Pedestrian Navigation," in *Journal of Global Positioning Systems*, vol. 6, 2007, pp. 56–64.

[117] B. Ristic and S. Arulampalam, *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artech House, 2004.

[118] T. Roosta, M. Wainwright, and S. Sastry, "Convergence analysis of reweighted sum-product algorithms", *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. II.541-544, 2007.

[119] R. Roy, T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques", *IEEE Trans. Signal Processing*, vol. 37, no. 7, July 1989.

[120] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis", *Ann. Math. Stat.*, vol. 24, no. 2, pp. 220-238, 1953.

[121] A. Sahai and D. Cabric, "Spectrum sensing: fundamental limits and practical challenges", *Proc. IEEE International Symp. New Frontiers Dynamic Spectrum Access Networks (DySPAN)*, Baltimore, MD, Nov. 2005.

[122] V. Saligrama, M. Alanyali, O. Savas, "Distributed Detection in Sensor Networks with Packet Losses and Finite Capacity Links", *IEEE Trans. on Signal Processing*, vol. 54, no. 11, Nov. 2006.

[123] V. Savic, A. Poblacion, S. Zazo, and M. Garcia, "Indoor positioning using nonparametric belief propagation based on spanning trees", *EURASIP Journal on Wireless Communications and Networking*, Aug. 2010.

[124] V. Savic, H. Wymeersch, F. Penna, S. Zazo, "Optimized Edge Appearance Probabilities for Cooperative Localization based on Non-parametric Tree-Reweighted Belief Propagation", *IEEE ICASSP*, Prague, Czech Republic, May 2011.

[125] R. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, Mar. 1986.

[126] J. Schott, "A note on the critical values used in stepwise tests for multiplicative components of interaction", *Comm. Stat. Th. Meth.*, vol 15, no. 5, pp. 1561-1570, 1986.

[127] Y. Sheikh and M. Shah, "Bayesian object detection in dynamic scenes", in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, June 2005

[128] S. J. Shellhammer, "Spectrum Sensing in IEEE 802.22", *EURASIP IAPR Workshop on Cognitive Information Processing*, June 2008.

[129] A. Soshnikov, "A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices", *J. Statist. Phys.* vol. 108, no. 5-6, pp. 1033–1056, 2002.

[130] A. Taherpour, M. Nasiri-Kenari, S. Gazor, "Multiple Antenna Spectrum Sensing in Cognitive Radios", *IEEE Trans. on Wireless Communications*, vol. 9, no. 2, Feb. 2010.

[131] R. Tandra, A. Sahai, "SNR Walls for Signal Detection", *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, Feb. 2008.

[132] R. R. Tenney, N. R. Sandell Jr., "Detection with distributed sensors", *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, July 1981.

[133] Texas Instruments, Chipcon CC2420 RF transceiver datasheet, http://focus.ti.com/lit/ds/symlink/cc2420.pdf

[134] TinyOS documentation, http://docs.tinyos.net/index.php/Main_Page

[135] R. Tomasi, H. Khaleel, F. Penna, C. Pastrone, M. Spirito, R. Garello, "Frequency Agility in IPv6-Based Wireless Personal Area Networks (6LoWPAN)" (invited paper), *Wired/Wireless Internet Communications (WWIC)*, Lulea, Sweden, June 1-3, 2010, in *Springer-Verlag Lecture Notes in Computer Science*, vol. 6074/2010, pp. 146-157

[136] C. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles", *Comm. Math. Phys*, vol.177, pp.727-754, 1996.

[137] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications", in *Foundations and Trends in Communications and Information Theory*, vol. 1, (Hanover, MA, USA), Now Publishers Inc., 2004.

[138] H. Urkowitz, "Energy Detection of Unknown Deterministic Signals", *Proc. of the IEEE*, vol. 55, no. 4, pp. 523-531, April 1967.

[139] O. Vermesan, M. Harrison et al., "Internet of Things Strategic Research Roadmap," *Technical report, Cluster of European Research Projects on the Internet of Things (CERP-IoT)*, Brussels, Belgium, September 2009.

[140] S. Venkatesh and M. Alanyali, "M-ary hypothesis testing in sensor networks", in *Proc. 38th Ann. Conf. Inform. Sci. Syst.*, Mar. 2004.

[141] R. Viswanathan, P. K. Varshney, "Distributed Detection with Multiple Sensors: Part I – Fundamentals", *Proc. of the IEEE*, vol. 85, no. 1, Jan. 1997.

[142] M. Wainwright and M. Jordan, "Graphical models, exponential families, and variational inference", *Foundations and Trends in Machine Learning*, vol. 1, no. 1-2, 2008.

[143] M. Wainwright, T. Jaakkola, A. Willsky, "A new class of upper bounds on the log partition function", *IEEE Transactions on Information Theory*, vol. 51, no. 7, p. 2313, 2005.

[144] W. Wang, M. Chatterjee, K. Kwiat, "Attacker Detection Game in Wireless Networks with Channel Uncertainty", *Proc. of IEEE ICC*, Cape Town, South Africa, June 2010.

[145] W. Wang, H. Li, Y. Sun, Z. Han, "Attack-proof collaborative spectrum sensing in cognitive radio networks", *Proc. 43rd Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD (USA), March 2009.

[146] W. Wang, H. Li, Y. Sun, Z. Han, "CatchIt: Detect Malicious Nodes in Collaborative Spectrum Sensing", *Proc. IEEE GLOBECOM*, Honolulu, HI (USA), Dec. 2009.

[147] W. Wang, Y. Sun, H. Li, Z. Han, "Cross-layer Attack and Defense in Cognitive Radio Networks", *Proc. IEEE GLOBECOM*, Miami, FL (USA), Dec. 2010.

[148] W. Wang, H. Li, Y. Sun, Z. Han, "Securing Collaborative Spectrum Sensing against Untrustworthy Secondary Users in Cognitive Radio Networks" *EURASIP Journal on Advances in Signal Processing*, Vol. 2010, Article ID 695750.

[149] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-33, no. 2, pp.387-392, Apr. 1985.

[150] L. Wei, O. Tirkkonen, "Cooperative spectrum sensing of OFDM signals using largest eigenvalue distributions", *Proc. IEEE PIMRC*, Tokyo, Japan, Sept. 2009.

[151] L. Wei, O. Tirkkonen, "Approximation to the Condition Number Distribution of Almost Square Matrices", *Proc. International Conference on Ultra-Modern Telecommunications (ICUMT 2009)*, St. Petersburg, Russia, Oct. 2009.

[152] Y. Weiss, "Correctness of local probability propagation in graphical models with loops" *Neural Computation*, vol. 12, 2000.

[153] N. Wiberg, "Codes and decoding on general graphs", PhD thesis, Linköping University, Sweden, 1996.

[154] A.S. Willsky, H.L. Jones, "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems", *IEEE Trans. on Automatic Control*, Feb. 1976.

[155] J. Wishart "The generalized product moment distribution in samples from a normal multivariate population", *Biometrika*, vol. 20A, pp. 32-52, 1928.

[156] A.P. Worthen and W.E. Stark, "Unified design of iterative receivers using factor graphs", *IEEE Transactions on Information Theory*, vol. 47, pp. 843-849, Feb. 2001.

[157] H. Wymeersch, *Iterative Receiver Design*, Cambridge University Press, 2007.

[158] H. Wymeersch, J. Lien, and M.Z. Win, "Cooperative localization in wireless networks," in *Proceedings of the IEEE*, vol. 97, no. 2, Feb. 2009, pp. 427–450.

[159] H. Wymeersch, F. Penna, V. Savic, "Uniformly Reweighted Belief Propagation for Distributed Inference", submitted to *IEEE Transactions on Wireless Communications*, 2011.

[160] H. Wymeersch, F. Penna, V. Savic, "Uniformly Reweighted Belief Propagation: A Factor Graph Approach", *IEEE International Symposium on Information Theory (ISIT)*, Saint Petersburg, Russia, July 2011.

[161] M. Xiang, J. Zhao, "On the Performance of Distributed Neyman-Pearson Detection Systems", *IEEE Trans. on System, Man, and Cybernetics*, vol. 31, no. 1, Jan. 2001.

[162] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms", *IEEE Transactions on Information Theory*, vol. 51, no. 7, 2005.

[163] S. Zarrin, T.J. Lim, "Belief Propagation on Factor Graphs for Cooperative Spectrum Sensing in Cognitive Radio", *IEEE DySPAN*, Chicago, IL (USA), Oct. 2008.

[164] S. Zarrin, T.J. Lim, "Cooperative Spectrum Sensing in Cognitive Radios With Incomplete Likelihood Functions", *IEEE Transactions on Signal Processing*, vol. 58, no. 6, June 2010.

[165] K. Zeng, P. Pawelczak, D. Cabric, "Reputation-based Cooperative Spectrum Sensing with Trusted Node Assistance", *IEEE Comm. Letters*, vol. 14, no. 3, March 2010.

[166] Y. H. Zeng and Y.-C. Liang, "Eigenvalue based spectrum sensing algorithms for cognitive radio", *IEEE Trans. on Communications*, vol. 57, no. 6, pp. 1784-1793, June 2009.

[167] Y. Zeng, Y.-C. Liang, "Maximum-Minimum Eigenvalue Detection for Cognitive Radio", *Proc. 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC) '07*, 2007.

[168] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A Review on Spectrum Sensing for Cognitive Radio: Challenges and Solutions", *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1-15, January 2010.

[169] Y. Zeng, Y.C. Liang, R. Zhang, "Blindly combined energy detection for spectrum sensing in cognitive radio", *IEEE Signal Processing Letters*, vol. 15, 2008.

[170] Q.T. Zhang, "Advanced Detection Techniques for Cognitive Radio", *Proc. of International Conference on Communications (ICC2009)*, 2009.

[171] R. Zhang, T.J. Lim, Y.C. Liang, Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: a GLRT approach", *IEEE Trans. Communications*, vol. 58, no. 1, Jan. 2010.

[172] Q. Zhu, Z. Han and T. Basar, "No-Regret Learning in Collaborative Spectrum Sensing with Malicious Nodes", *Proc. IEEE ICC 2010*, Cape Town, South Africa, June 2010.